Analisi bereizlea

Discriminant analysis

Y kualitatiboa:

m modalitate

Ω	X_1		X_j		X_p	**Y**
ω_1	x_{11}		x_{1j}		x_{1p}	y_1
ω_i	x_{i1}		x_{ij}		x_{ip}	y_i
ω_n	x_{n1}		x_{nj}		x_{np}	y_n

Analisi bereizlea

Discriminant analysis

Y kualitatiboa:
m modalitate

Ω	X_1		X_j		X_p	**Y**
ω_1	x_{11}		x_{1j}		x_{1p}	y_1
ω_i	x_{i1}		x_{ij}		x_{ip}	y_i
ω_n	x_{n1}		x_{nj}		x_{np}	y_n

Funtzio bereizleak
(modalitateek funtzio bana):

$$Y_1 : g_1(X_1, X_2, ..., X_p)$$

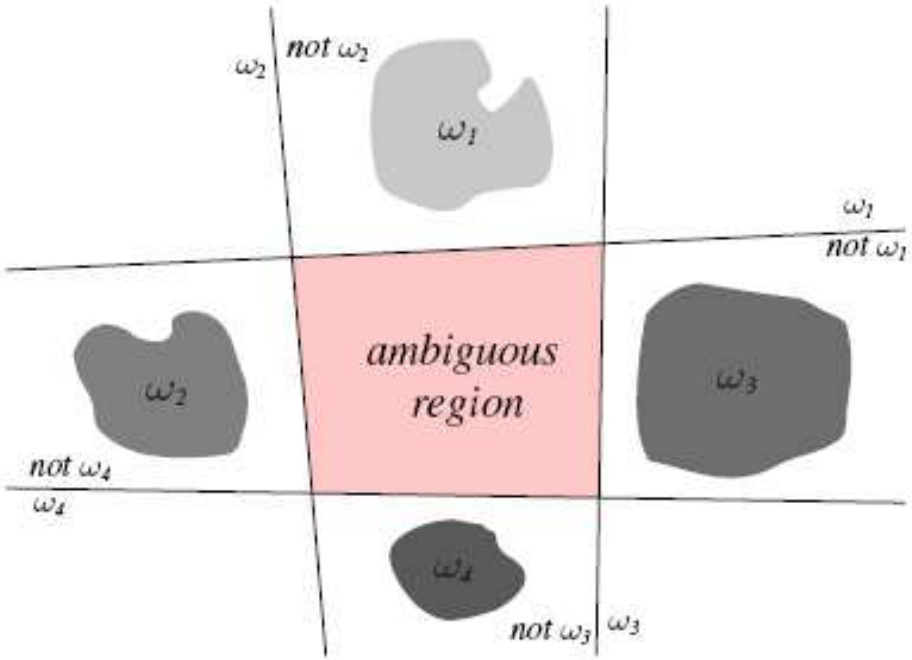
$$Y_2 : g_2(X_1, X_2, ..., X_p)$$

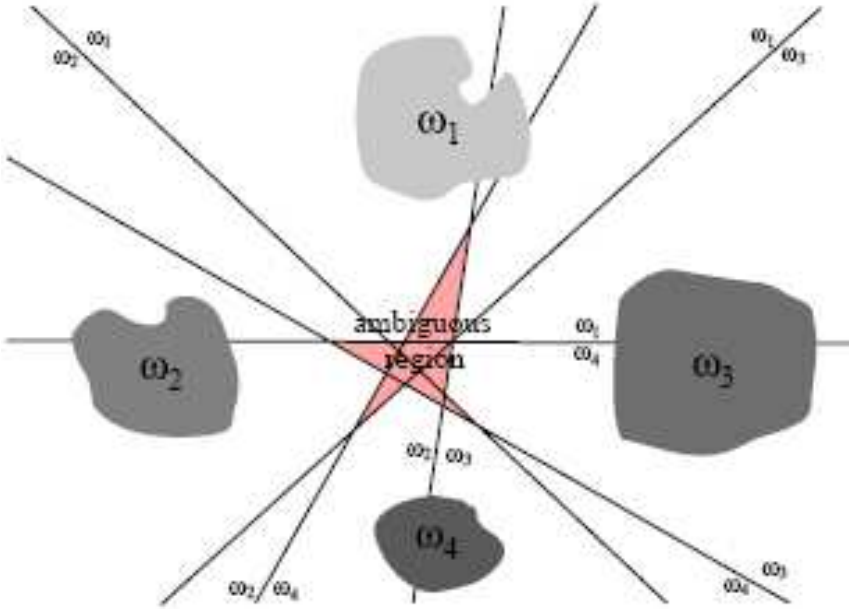
...

$$Y_m : g_m(X_1, X_2, ..., X_p)$$

Analisi bereizlea

Discriminant analysis





Analisi bereizlea

Discriminant analysis

Y kualitatiboa:
m modalitate

Ω	X_1		X_j		X_p	**Y**
ω_1	x_{11}		x_{1j}		x_{1p}	y_1
ω_i	x_{i1}		x_{ij}		x_{ip}	y_i
ω_n	x_{n1}		x_{nj}		x_{np}	y_n

Funtzio bereizleak
(modalitateek funtzio bana):

$$Y_1 : g_1(X_1, X_2, ..., X_p)$$

$$Y_2 : g_2(X_1, X_2, ..., X_p)$$

...

$$Y_m : g_m(X_1, X_2, ..., X_p)$$

Objektu bat $\boldsymbol{Y_k}$koa dela erabakitzen da, baldin bada:

$$g_k(X_1(\omega), X_2(\omega), ..., X_p(\omega)) = \max_{j=1}^{m} g_j(X_1(\omega), X_2(\omega), ..., X_p(\omega))$$

Analisi bereizle lineala

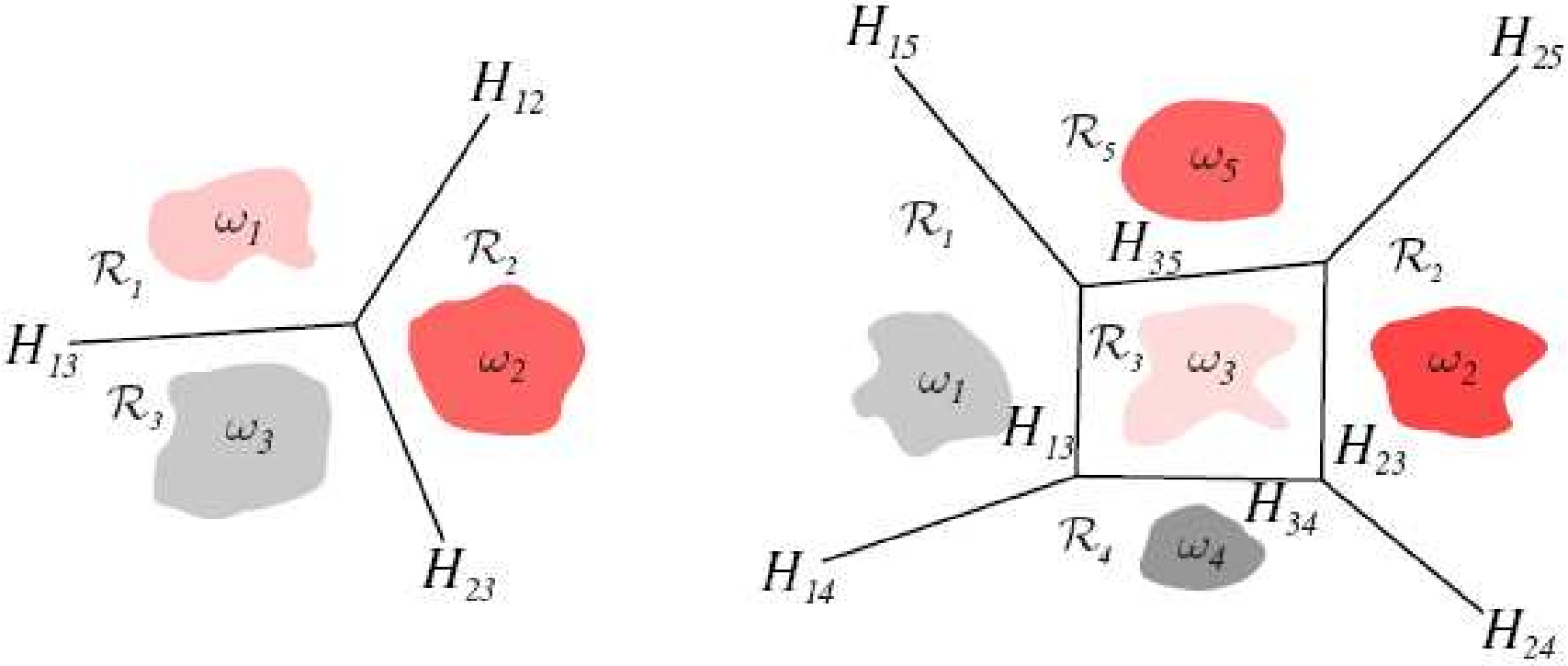
Linear Discriminant Analysis (LDA)

Objektu bat $\boldsymbol{Y_k}$koa dela erabakitzen da, baldin bada:

$$g_k(X_1(\omega), X_2(\omega), ..., X_p(\omega)) = \max_{j=1}^{m} g_j(X_1(\omega), X_2(\omega), ..., X_p(\omega))$$

(gogoratu *k*-auzokide gertuenen teknika, *K-NN*)

Ondorioz, errepresentazio-espazioa zatikatu egiten da



Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

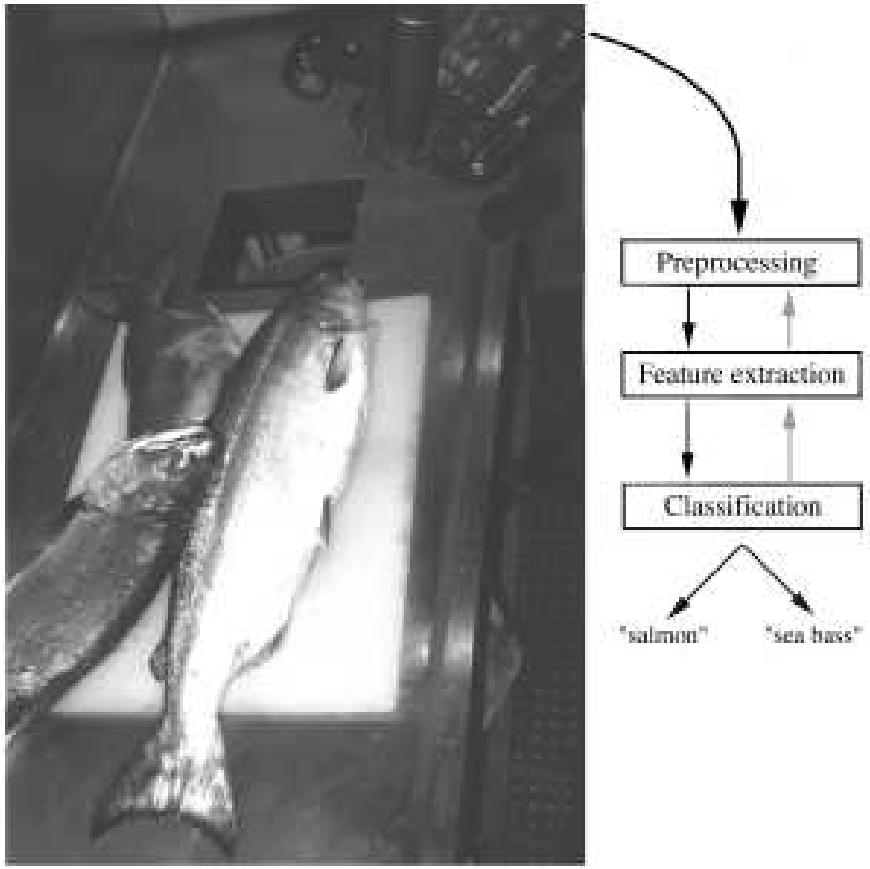


FIGURE 1.1. The objects to be classified are first sensed by a transducer (camera), whose signals are preprocessed. Next the features are extracted and finally the classification is emitted, here either "salmon" or "sea bass." Although the information flow is often chosen to be from the source to the classifier, some systems employ information flow in which earlier levels of processing can be altered based on the tentative or preliminary response in later levels (gray arrows). Yet others combine two or more stages into a unified step, such as simultaneous segmentation and feature extraction. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

X*ekiko *Y*ren *analisi bereizlea

Menpeko aldagaia kualitatibo bitarra

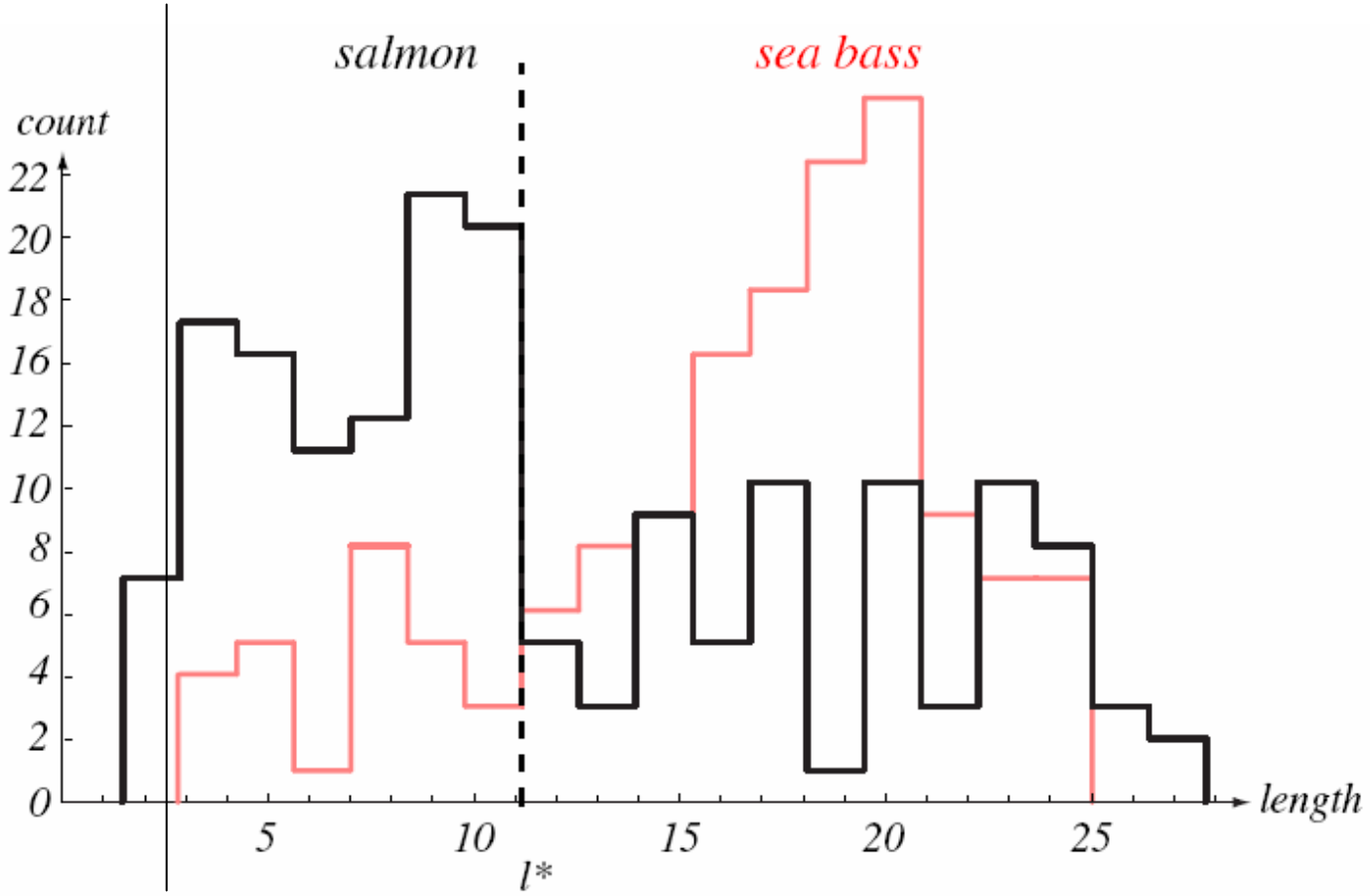
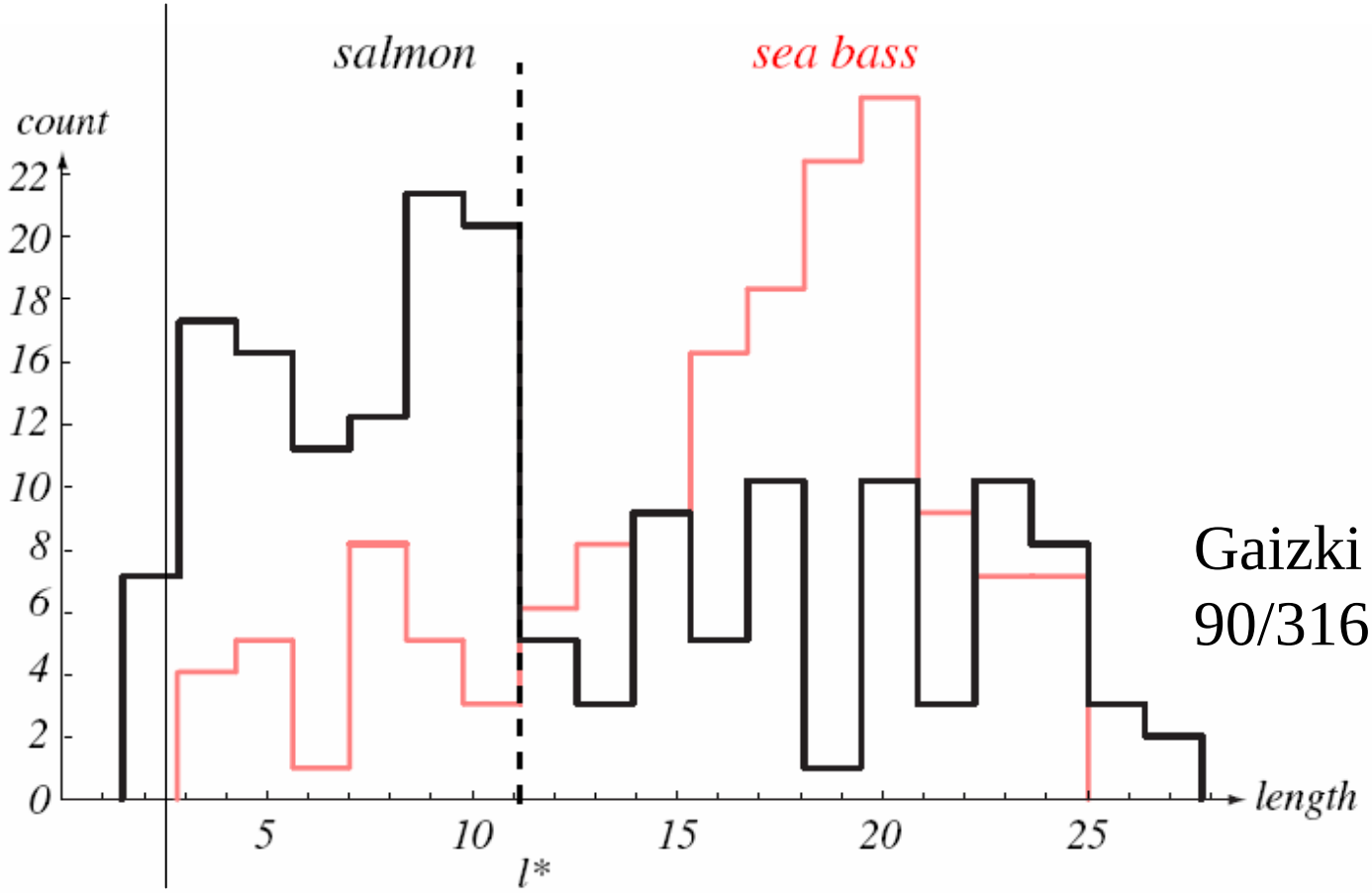


FIGURE 1.2. Histograms for the length feature for the two categories. No single threshold value of the length will serve to unambiguously discriminate between the two categories; using length alone, we will have some errors. The value marked *l** will lead to the smallest number of errors, on average. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

X*ekiko *Y*ren *analisi bereizlea

Menpeko aldagaia kualitatibo bitarra



Gaizki sailkatuak:
90/316=%28

FIGURE 1.2. Histograms for the length feature for the two categories. No single threshold value of the length will serve to unambiguously discriminate between the two categories; using length alone, we will have some errors. The value marked *l** will lead to the smallest number of errors, on average. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

X*ekiko *Y*ren *analisi bereizlea

Menpeko aldagaia kualitatibo bitarra

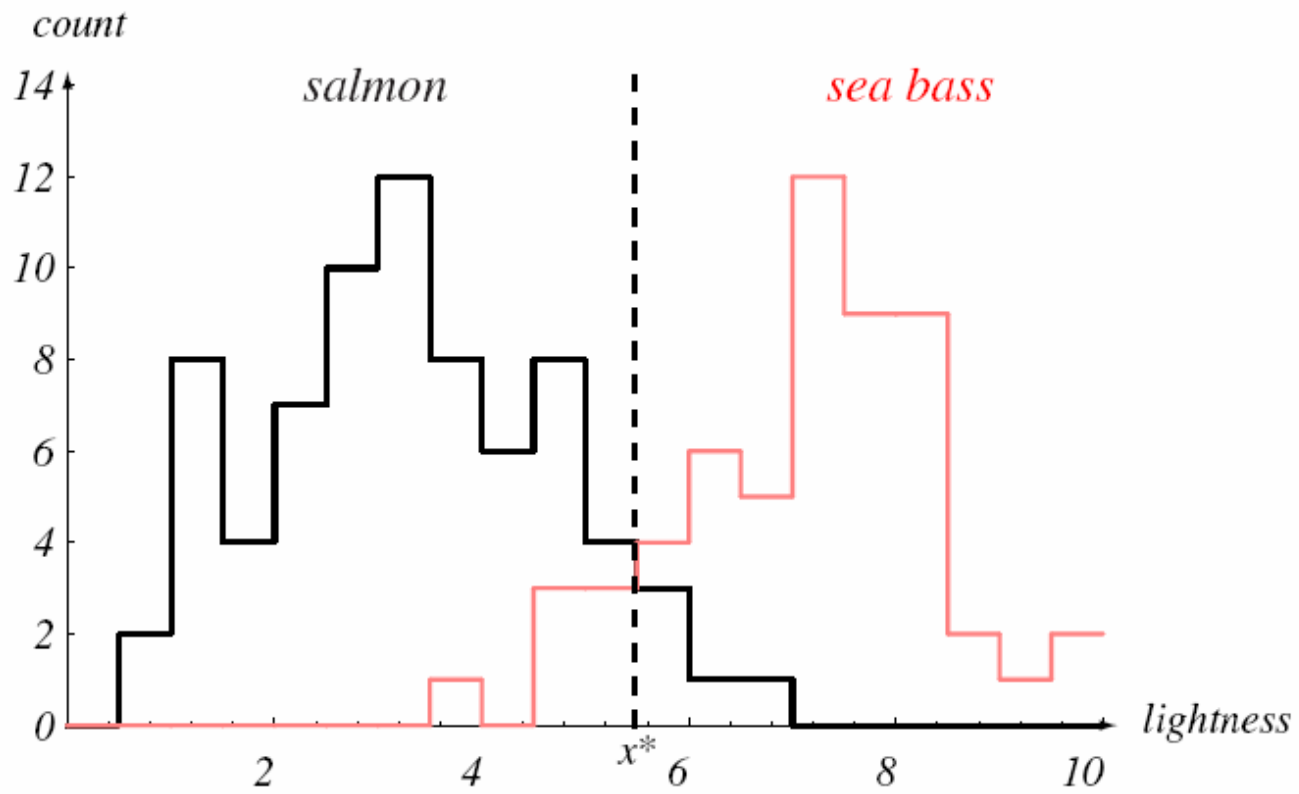
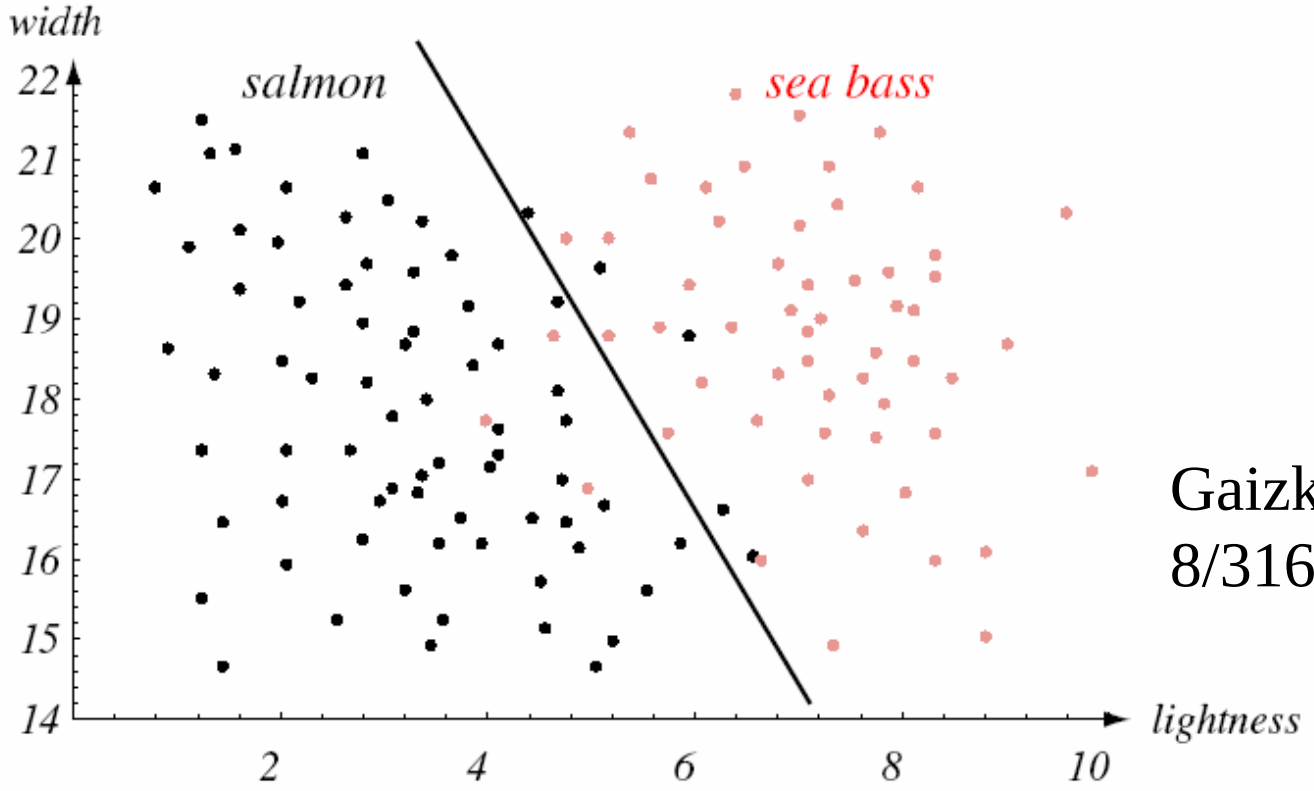


FIGURE 1.3. Histograms for the lightness feature for the two categories. No single threshold value x^* (decision boundary) will serve to unambiguously discriminate between the two categories; using lightness alone, we will have some errors. The value x^* marked will lead to the smallest number of errors, on average. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

X*ekiko *Y*ren *analisi bereizlea

Menpeko aldagaia kualitatibo bitarra



Gaizki sailkatuak:
8/316=%2.5

FIGURE 1.4. The two features of lightness and width for sea bass and salmon. The dark line could serve as a decision boundary of our classifier. Overall classification error on the data shown is lower than if we use only one feature as in Fig. 1.3, but there will still be some errors. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

X*ekiko *Y*ren *analisi bereizlea

Menpeko aldagaia kualitatibo bitarra

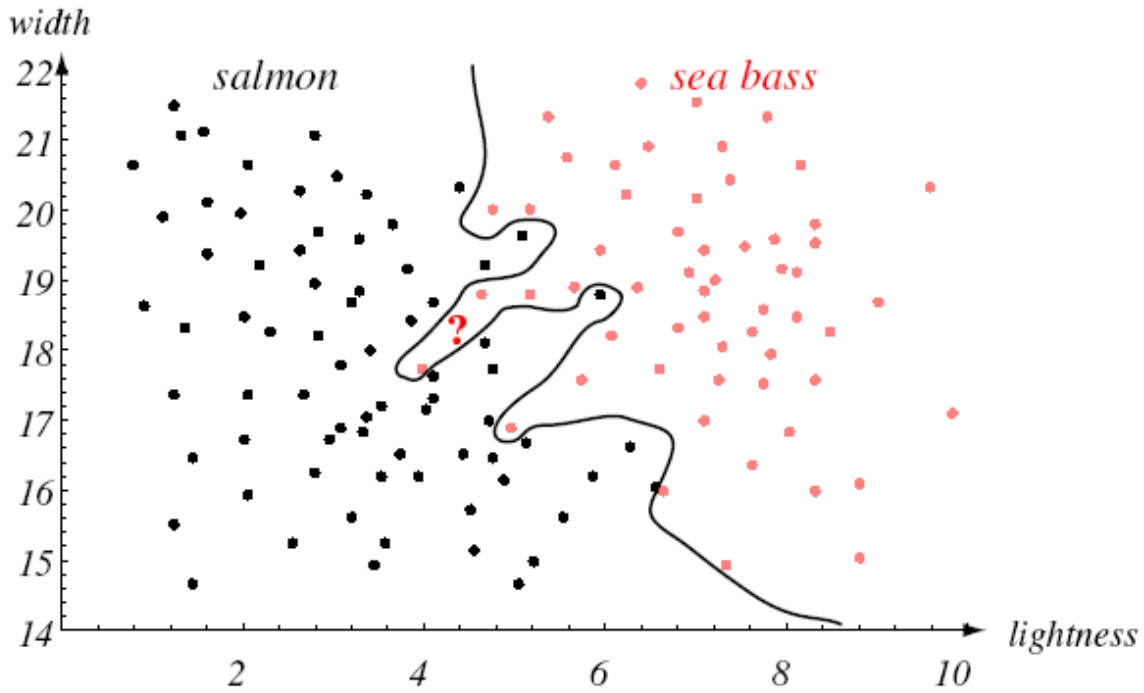


FIGURE 1.5. Overly complex models for the fish will lead to decision boundaries that are complicated. While such a decision may lead to perfect classification of our training samples, it would lead to poor performance on future patterns. The novel test point marked **?** is evidently most likely a salmon, whereas the complex decision boundary shown leads it to be classified as a sea bass. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

X*ekiko *Y*ren *analisi bereizlea

Menpeko aldagaia kualitatibo bitarra

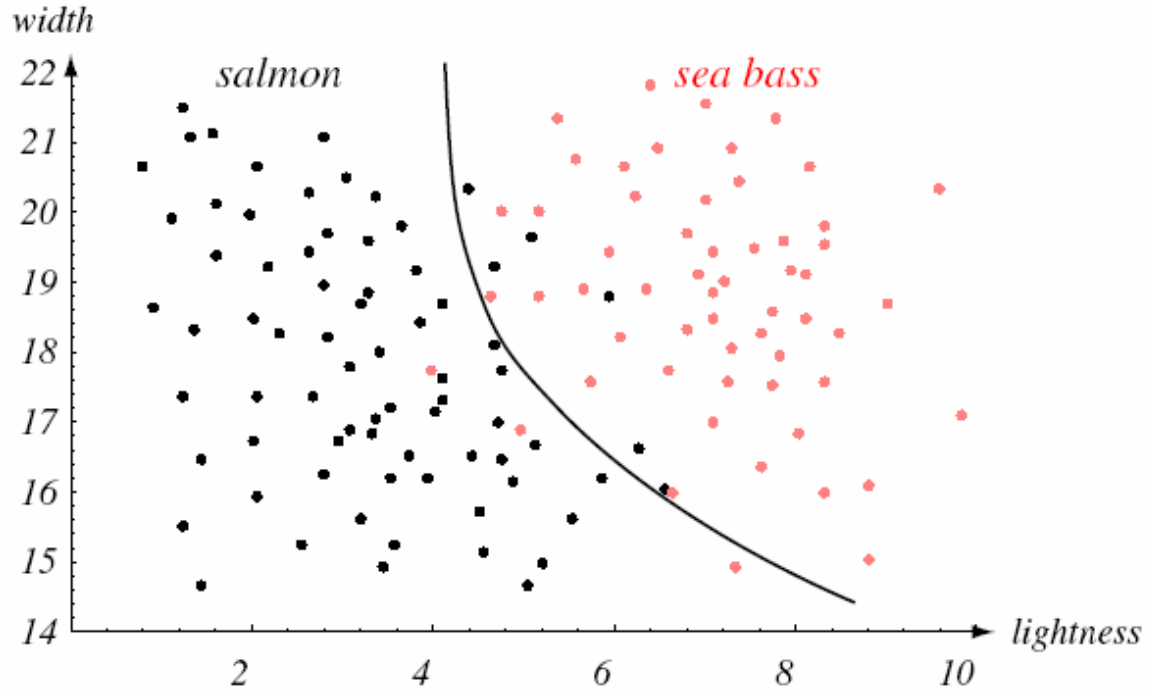


FIGURE 1.6. The decision boundary shown might represent the optimal tradeoff between performance on the training set and simplicity of classifier, thereby giving the highest accuracy on new patterns. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Analisi bereizlea

Discriminant analysis

$g_1, g_{2,}..., g_m$ funtzio *bereizleak* eraiki behar dira

$m=2$ (***Y*** aldagaia bitarra da)

$\boldsymbol{X_j}$ aldagaiak kuantitatiboak dira

Erabakitze-araua:

Baldin bada $g_1(X_1, X_2, ..., X_p) > g_2(X_1, X_2, ..., X_p)$,

orduan Y_1 ; bestela Y_2

$$g = g_1 - g_2$$

Baldin bada $g(X_1, X_1, ..., X_p) > 0$ orduan Y_1 ; bestela Y_2

Muga *bereizlea*: $g(X_1, X_1, ..., X_p) = 0$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

$g_1, g_2, \ldots, g_m$ funtzio *bereizleak* eraiki behar dira

$m=2$ (***Y*** aldagaia bitarra da)

$\boldsymbol{X_j}$ aldagaiak kuantitatiboak dira

g lineala da:

$$g_1(X_1, X_2, \ldots, X_p) = a_{10} + a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p$$

$$g_2(X_1, X_2, \ldots, X_p) = a_{20} + a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p$$

$$g = g_1 - g_2 \text{ lineala}$$

g lineala izan daiteke g_1 eta g_2 linealak izan gabe

$$g(X_1, X_2, \ldots, X_p) = a_0 + a_1X_1 + a_2X_2 + \ldots + a_pX_p$$

$g(X_1, X_2, \ldots, X_p) = 0$ *hiperplano* bat da.

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

g funtzio *bereizlea* eraiki behar da; hau da:

$a_0, a_1, a_2, ..., a_p$ koefizienteak

$a_0 + a_1 X_1 + a_2 X_2 + ... + a_p X_p = 0$ *hiperplano*ko bi puntu $\omega^{(1}$ eta $\omega^{(2}$:

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

g funtzio *bereizlea* eraiki behar da; hau da:

$a_0, a_1, a_2, ..., a_p$ koefizienteak

$a_0 + a_1 X_1 + a_2 X_2 + ... + a_p X_p = 0$ *hiperplano*ko bi puntu $\omega^{(1}$ eta $\omega^{(2}$:

$$a_0 + a_1 x_1^{(1} + a_2 x_2^{(1} + ... + a_p x_p^{(1} = a_0 + a_1 x_1^{(2} + a_2 x_2^{(2} + ... + a_p x_p^{(2}$$

$$a_1 (x_1^{(1} - x_1^{(2}) + a_2 (x_2^{(1} - x_2^{(2}) + ... + a_p (x_p^{(1} - x_p^{(2}) = 0$$

$(x^{(1} - x^{(2})$ bektorea hiperplanokoa da eta elkarzuta da *a*rekiko.

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

g funtzio *bereizlea* eraiki behar da; hau da:

$a_0, a_1, a_2, ..., a_p$ koefizienteak

$a_0 + a_1 X_1 + a_2 X_2 + ... + a_p X_p = 0$ *hiperplano*ko bi puntu $\omega^{(1}$ eta $\omega^{(2}$:

$$a_0 + a_1 x_1^{(1} + a_2 x_2^{(1} + ... + a_p x_p^{(1} = a_0 + a_1 x_1^{(2} + a_2 x_2^{(2} + ... + a_p x_p^{(2}$$

$$a_1 (x_1^{(1} - x_1^{(2}) + a_2 (x_2^{(1} - x_2^{(2}) + ... + a_p (x_p^{(1} - x_p^{(2}) = 0$$

$(x^{(1} - x^{(2})$ bektorea hiperplanokoa da eta elkarzuta da *a*rekiko.

x^* puntu baten distantzia H hiperplanoarekiko:

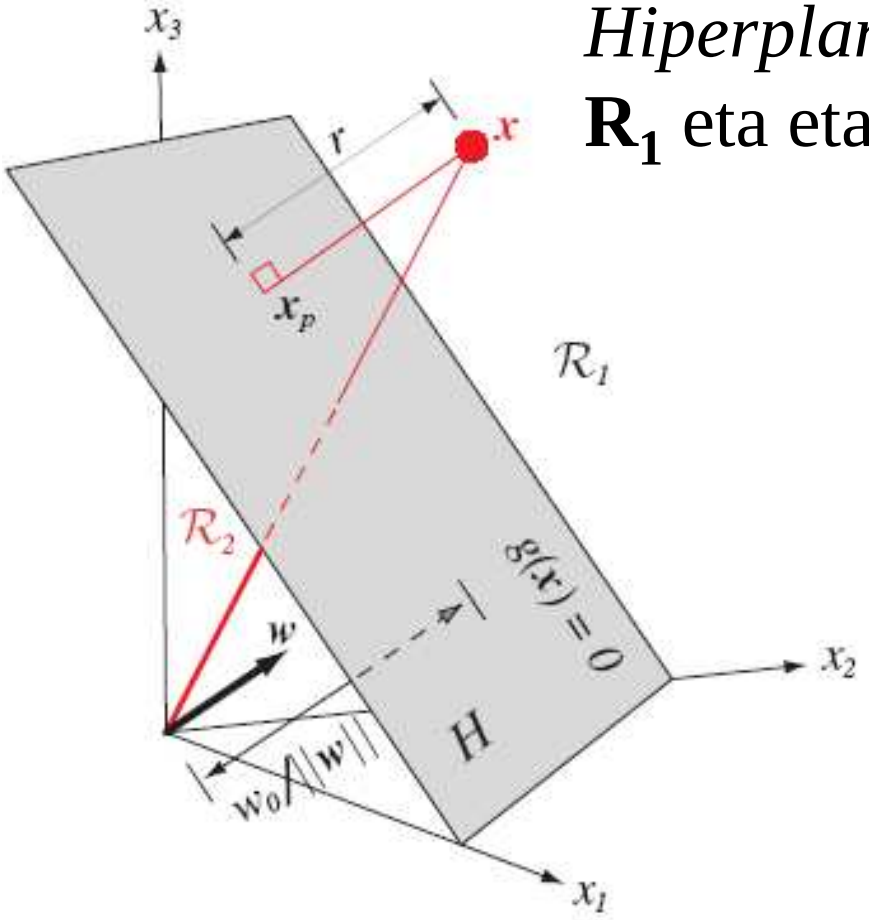
$$d(x^*, H) \equiv r = \frac{a_0 + a_1 x_1^* + a_2 x_2^* + ... + a_p x_p^*}{\sqrt{a_1^2 + a_2^2 + ... + a_p^2}}$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

g funtzio *bereizlea* eraiki behar da; hau da:

$a_0, a_1, a_2, ..., a_p$ koefizienteak



*Hiperplanoa*k bitan zatitzen du espazioa: $\mathbf{R_1}$ eta eta $\mathbf{R_2}$:

R_1 :

$g(x_1, x_2, ..., x_p) > 0$

$a_0 + a_1 x_1 + a_2 x_2 + ... + a_p x_p > 0$

R_2 :

$g(x_1, x_2, ..., x_p) < 0$

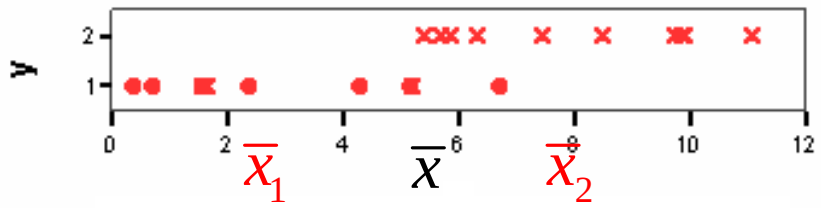
$a_0 + a_1 x_1 + a_2 x_2 + ... + a_p x_p < 0$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



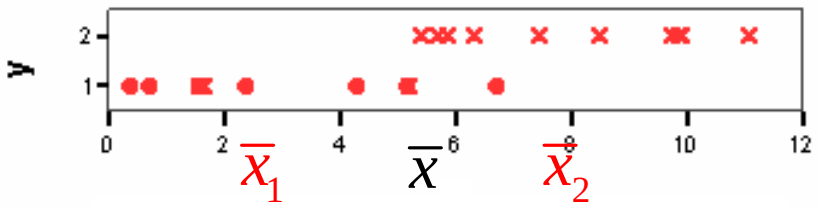
Y		Batezbest.	Desb.tip.
1	X	2,9474722	2,182214
2	X	7,9758592	2,095145
Total	X	5,4616657	3,314947

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



Y		Batezbest.	Desb.tip.
1	X	2,9474722	2,182214
2	X	7,9758592	2,095145
Total	X	5,4616657	3,314947

Funtzio bereizleak:

$$g_1(x) = x - \overline{x}_1$$

$$g_2(x) = x - \overline{x}_2$$

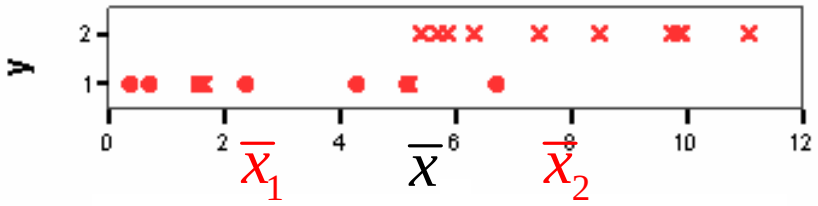
$$g(x) = g_1(x) - g_2(x) = \overline{x}_2 - \overline{x}_1$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



Y		Batezbest.	Desb.tip.
1	X	2,9474722	2,182214
2	X	7,9758592	2,095145
Total	X	5,4616657	3,314947

Funtzio bereizleak:

$$g_1(x) = x - \overline{x}_1$$

$$g_2(x) = x - \overline{x}_2$$

$$g(x) = g_1(x) - g_2(x) = \overline{x}_2 - \overline{x}_1$$

Erabakitze-araua:

$g(x) > 0$ denez, edozein objektu **2**kotzat hartu

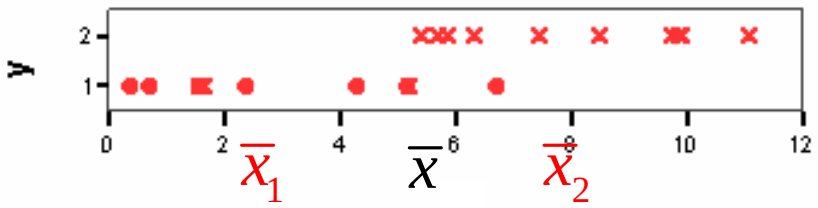
!!!!

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



Funtzio bereizleak:

$$g_1(x) = \left| x - \overline{x}_1 \right|$$

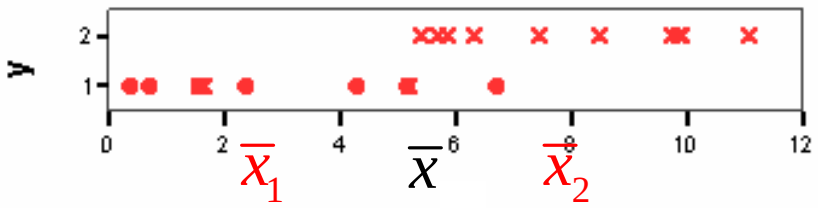
$$g_2(x) = \left| x - \overline{x}_2 \right|$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



Funtzio bereizleak:

$$g_1(x) = |x - \overline{x}_1|$$

$$g_2(x) = |x - \overline{x}_2|$$

$$g(x) = \overline{x}_2 - \overline{x}_1 \quad < 0 \quad x < \overline{x}_1$$

$$g(x) = 2\left(x - \frac{\overline{x}_1 + \overline{x}_2}{2}\right) \quad < 0 \quad x < \frac{\overline{x}_1 + \overline{x}_2}{2}$$

$$g(x) = 2\left(x - \frac{\overline{x}_1 + \overline{x}_2}{2}\right) \quad > 0 \quad x < \frac{\overline{x}_1 + \overline{x}_2}{2}$$

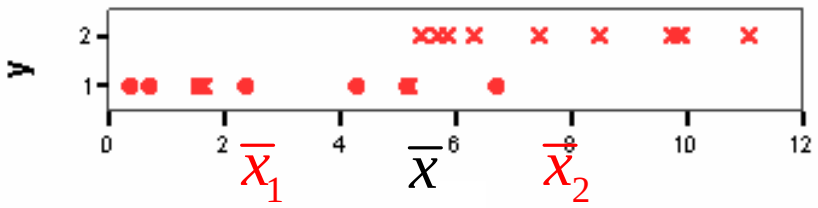
$$g(x) = \overline{x}_2 - \overline{x}_1 \quad > 0 \quad x > \overline{x}_2$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



Funtzio bereizleak:

$$g_1(x) = |x - \overline{x}_1|$$

$$g_2(x) = |x - \overline{x}_2|$$

$$g(x) = \overline{x}_2 - \overline{x}_1 \quad < 0 \quad x < \overline{x}_1$$

$$g(x) = 2\left(x - \frac{\overline{x}_1 + \overline{x}_2}{2}\right) \quad < 0 \quad x < \frac{\overline{x}_1 + \overline{x}_2}{2}$$

$\mathbf{R_1}$

$$g(x) = 2\left(x - \frac{\overline{x}_1 + \overline{x}_2}{2}\right) \quad > 0 \quad x < \frac{\overline{x}_1 + \overline{x}_2}{2}$$

$$g(x) = \overline{x}_2 - \overline{x}_1 \quad > 0 \quad x > \overline{x}_2$$

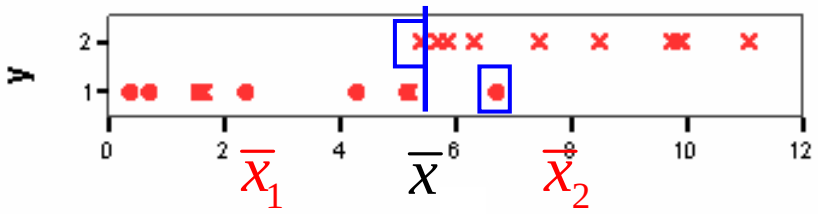
$\mathbf{R_2}$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



$$\frac{\overline{x}_1 + \overline{x}_2}{2} = 5.4617 \cong \overline{x}$$

Funtzio bereizleak:

$$g_1(x) = |x - \overline{x}_1|$$

$$g_2(x) = |x - \overline{x}_2|$$

$$g(x) = \overline{x}_2 - \overline{x}_1 \quad < 0 \quad x < \overline{x}_1$$

$$g(x) = 2\left(x - \frac{\overline{x}_1 + \overline{x}_2}{2}\right) \quad < 0 \quad x < \frac{\overline{x}_1 + \overline{x}_2}{2}$$

$\mathbf{R_1}$

$$g(x) = 2\left(x - \frac{\overline{x}_1 + \overline{x}_2}{2}\right) \quad > 0 \quad x < \frac{\overline{x}_1 + \overline{x}_2}{2}$$

$$g(x) = \overline{x}_2 - \overline{x}_1 \quad > 0 \quad x > \overline{x}_2$$

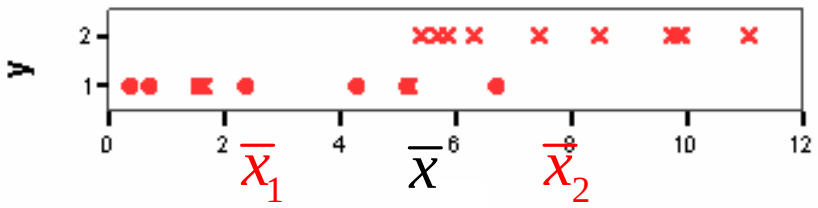
$\mathbf{R_2}$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



Funtzio bereizleak:

$$g_1(x) = \left(x - \overline{x}_1\right)^2$$

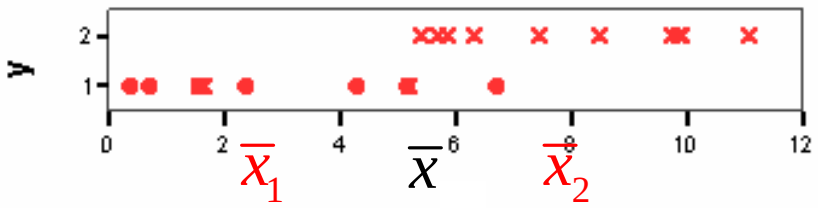
$$g_2(x) = \left(x - \overline{x}_2\right)^2$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



Funtzio bereizleak:

$$g_1(x) = \left(x - \overline{x}_1\right)^2$$

$$g_2(x) = \left(x - \overline{x}_2\right)^2$$

$$g(x) = 2\left(\overline{x}_2 - \overline{x}_1\right)\left(x - \frac{\overline{x}_1 + \overline{x}_2}{2}\right)$$

$$<0 \qquad x < \frac{\overline{x}_1 + \overline{x}_2}{2}$$

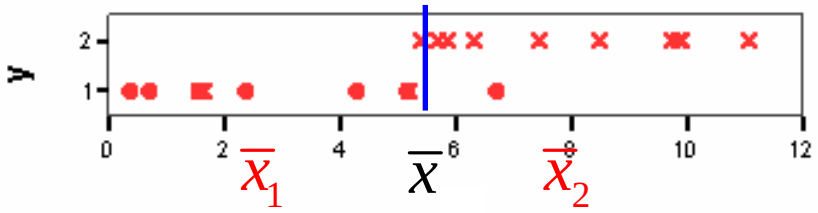
$$>0 \qquad x > \frac{\overline{x}_1 + \overline{x}_2}{2}$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



Funtzio bereizleak:

$$g_1(x) = \left(x - \overline{x}_1\right)^2$$

$$g_2(x) = \left(x - \overline{x}_2\right)^2$$

$$g(x) = 2\left(\overline{x}_2 - \overline{x}_1\right)\left(x - \frac{\overline{x}_1 + \overline{x}_2}{2}\right)$$

R_1

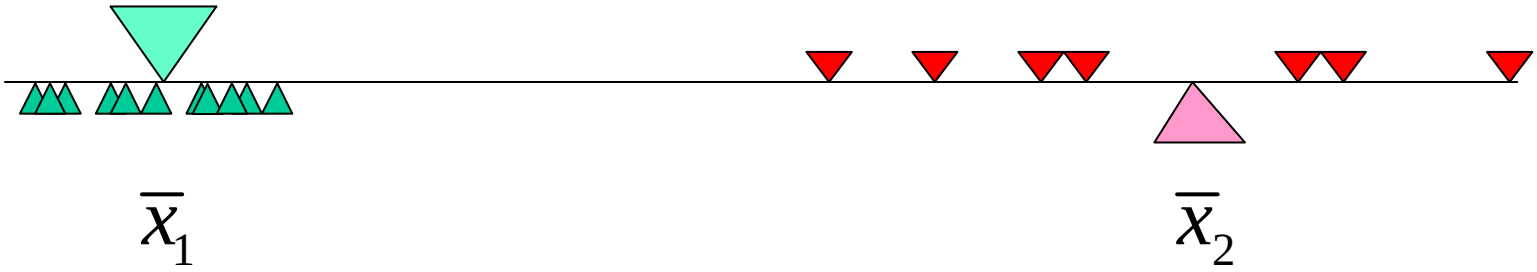
$$<0 \qquad x < \frac{\overline{x}_1 + \overline{x}_2}{2}$$

$$>0 \qquad x > \frac{\overline{x}_1 + \overline{x}_2}{2}$$

R_2

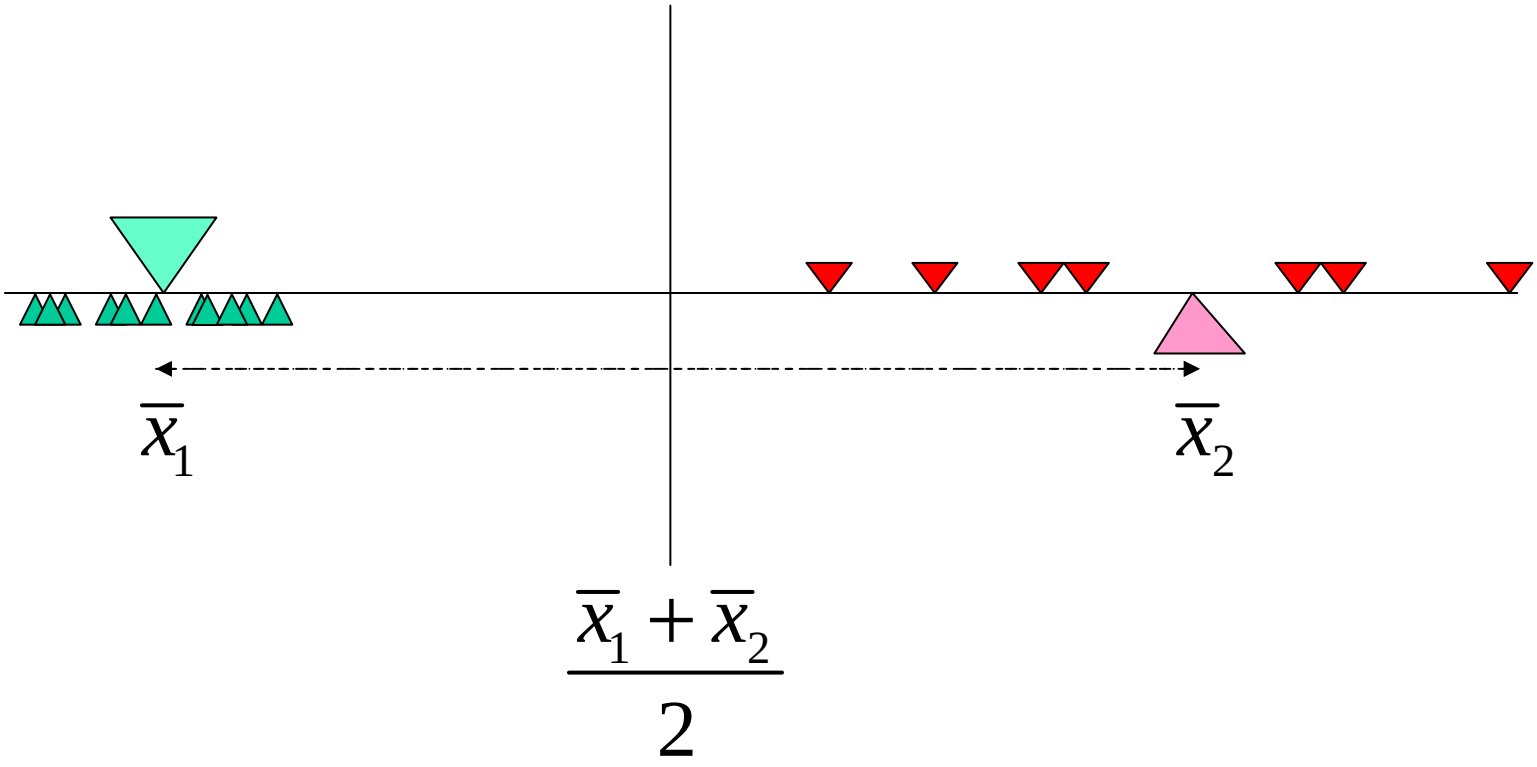
Analisi bereizle lineala

Linear Discriminant Analysis (LDA)



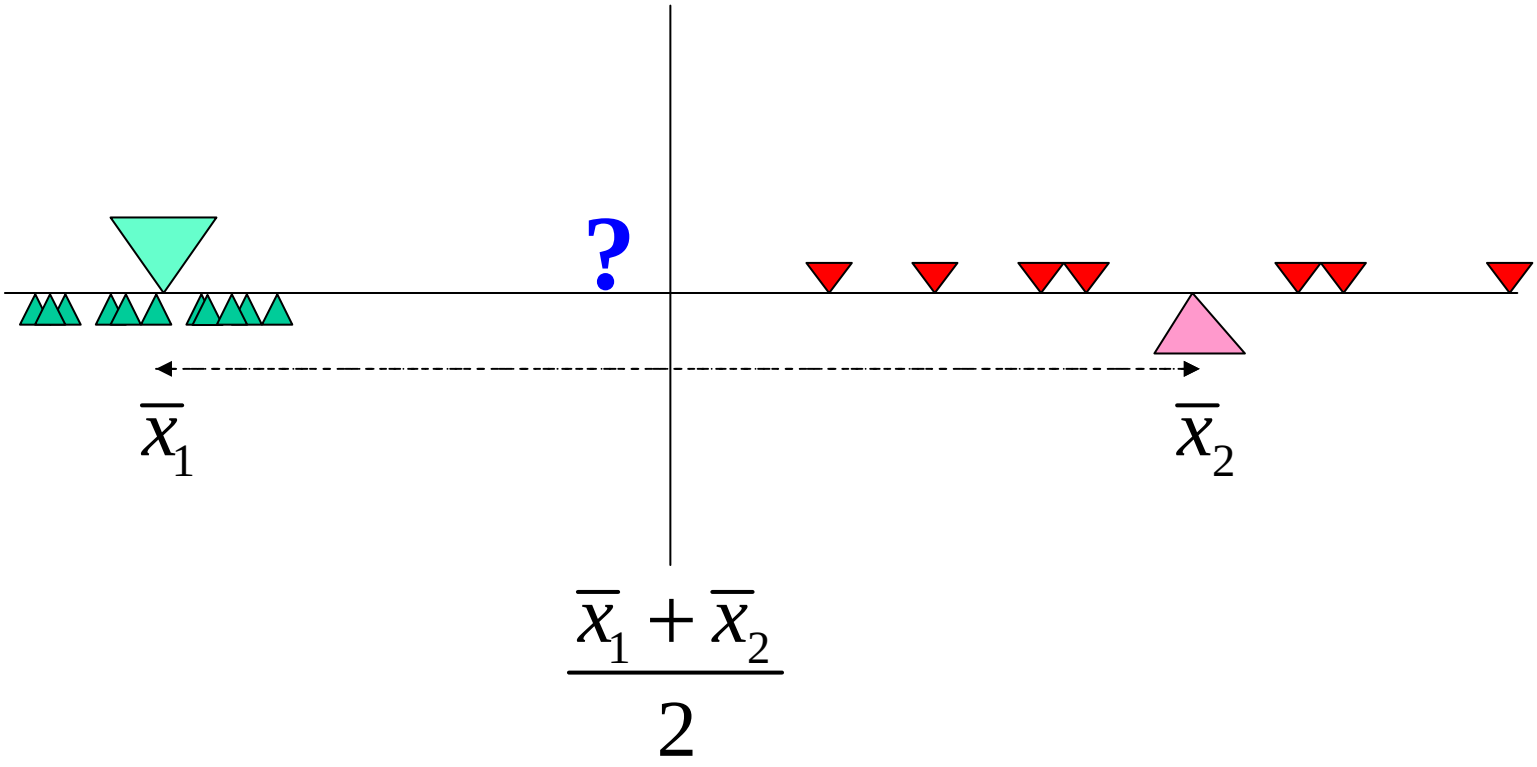
Analisi bereizle lineala

Linear Discriminant Analysis (LDA)



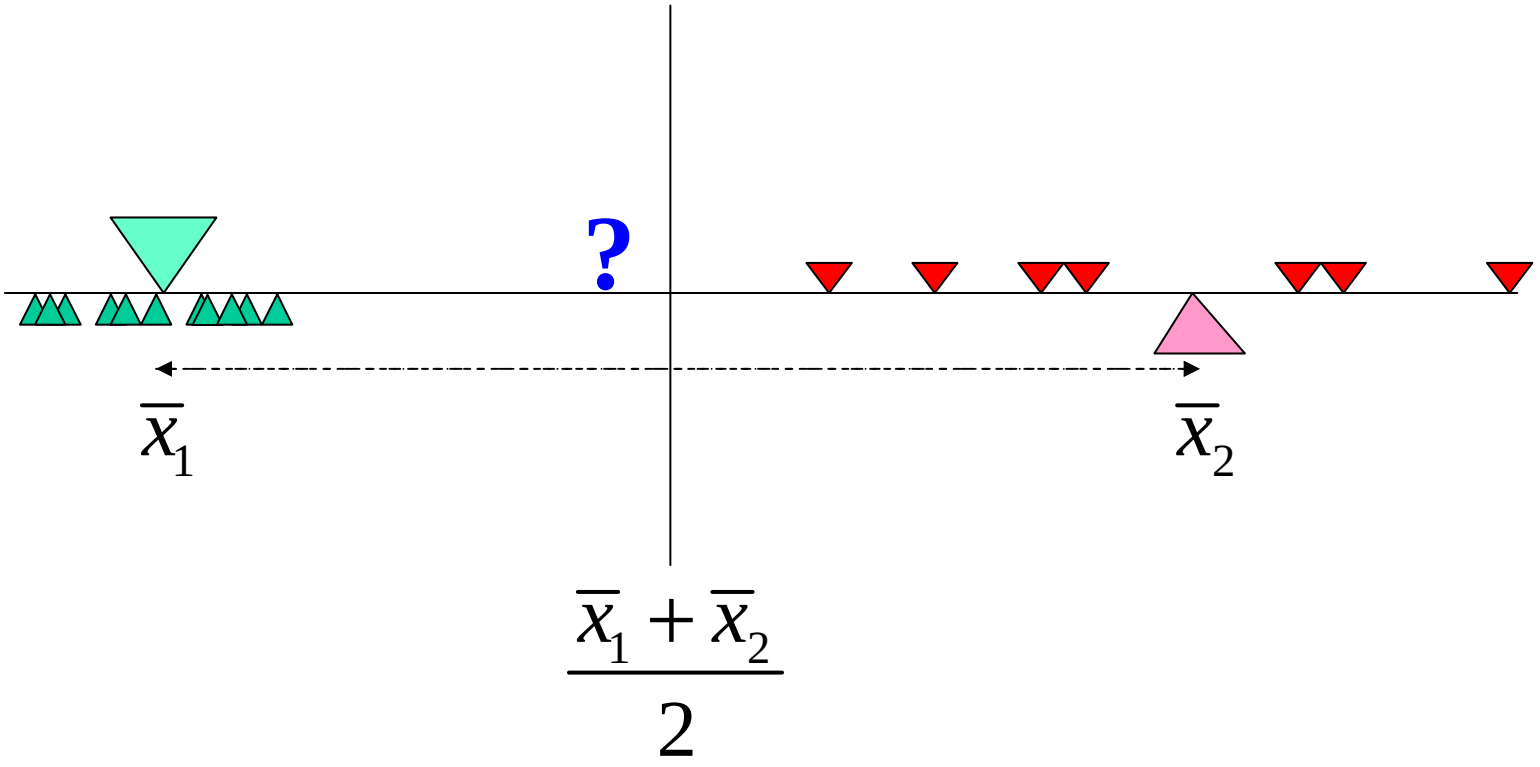
Analisi bereizle lineala

Linear Discriminant Analysis (LDA)



Analisi bereizle lineala

Linear Discriminant Analysis (LDA)



Senak eskatzen du ? ▼ gisa sailkatzea, baina arauak dio ▽

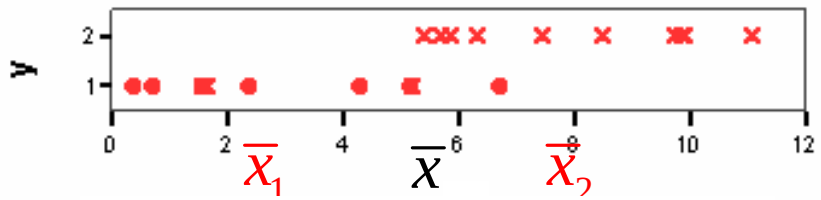
Zer dela eta? Sailen sakabanatzea: ***bariantzak*** kontutan hartu

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



Y		Media	Desv. típ.
1	X	2,9474722	2,182214
2	X	7,9758592	2,095145
Total	X	5,4616657	3,314947

Funtzio bereizleak:

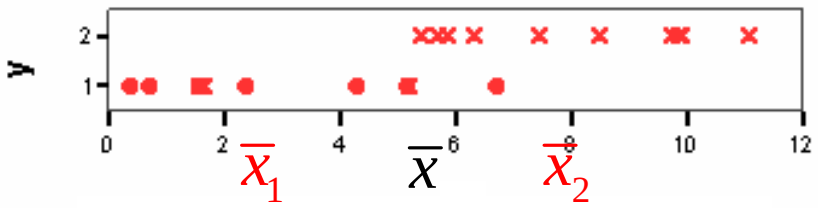
$$g_1(x) = \left(\frac{x - \overline{x}_1}{s_1} \right)^2 \qquad g_2(x) = \left(\frac{x - \overline{x}_2}{s_2} \right)^2$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

ident	x	y
w1	,70396	1
w2	5,12851	1
w3	1,57788	1
w4	1,54994	1
w5	4,27062	1
w6	6,67777	1
w7	1,65831	1
w8	2,36962	1
w9	,36968	1
w10	5,16843	1
w11	5,83867	2
w12	5,41226	2
w13	6,31614	2
w14	5,69690	2
w15	9,73109	2
w16	9,84856	2
w17	11,0675	2
w18	7,45862	2
w19	8,47496	2
w20	9,91385	2



Y		Media	Desv. típ.
1	X	2,9474722	2,182214
2	X	7,9758592	2,095145
Total	X	5,4616657	3,314947

Funtzio bereizleak:

$$g_1(x) = \left(\frac{x - \overline{x}_1}{s_1} \right)^2 \qquad g_2(x) = \left(\frac{x - \overline{x}_2}{s_2} \right)^2$$

$$g(x) = \frac{(s_2 - s_1)x^2 + 2x(\overline{x}_2 - \overline{x}_1) + (s_1\overline{x}_2^2 - s_2\overline{x}_1^2)}{s_1^2 s_2^2}$$

Ez da lineala

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

g funtzio *bereizlea* eraiki behar da; hau da:

$a_0, a_1, a_2, ..., a_p$ koefizienteak

Funtzioak nolanahi definiti baino, optimizazio-problema batez

Analisi bereizle lineala

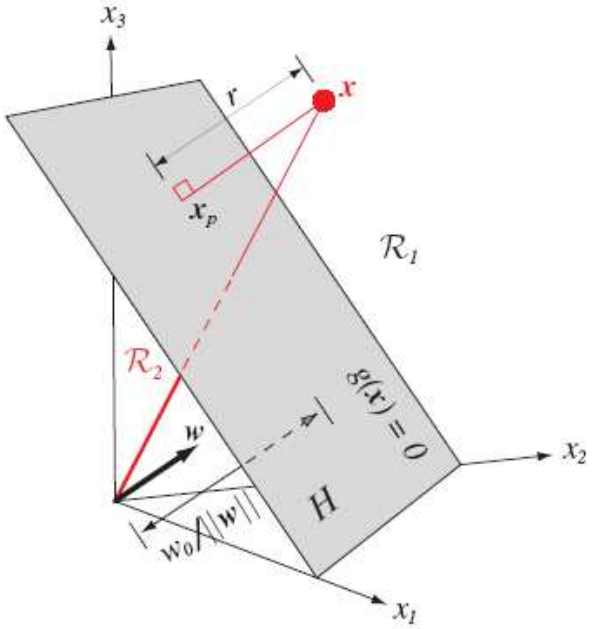
Linear Discriminant Analysis (LDA)

g funtzio *bereizlea* eraiki behar da; hau da:

$a_0, a_1, a_2, ..., a_p$ koefizienteak

Funtzioak nolanahi definiti baino, optimizazio-problema batez

Yren modalitate batekoei *+1* egokitu
(hiperplanoaren alde batean), eta besteari *-1*



Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

g funtzio *bereizlea* eraiki behar da; hau da:

$a_0, a_1, a_2, ..., a_p$ koefizienteak

Funtzioak nolanahi definiti baino, optimizazio-problema batez

Yren modalitate batekoei *+1* egokitu
(hiperplanoaren alde batean), eta besteari *-1*

Erregresio linealaren problema bera azaldu:

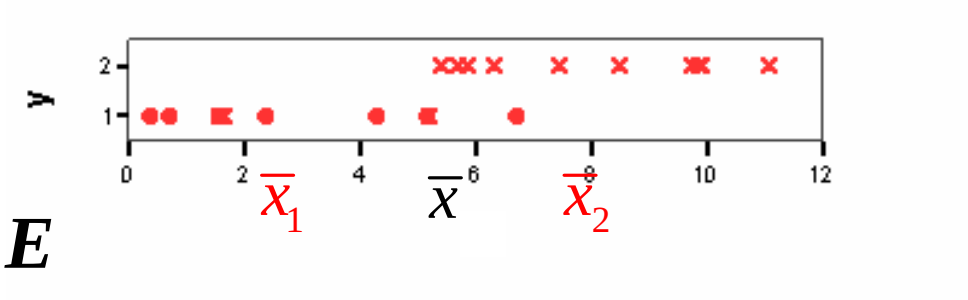
$$\boldsymbol{Y = a_0 + a_1 X_1 + a_2 X_2 + ... a_j X_j + ... + a_p X_p + E}$$

$$Y(\omega_i) = a_0 + a_1 X_1(\omega_i) + a_2 X_2(\omega_i) + ... a_j X_j(\omega_i) + ... + a_p X_p(\omega_i) + E(\omega_i)$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:

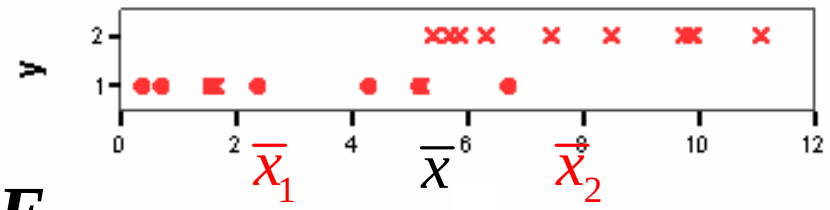


$$Y = b + a \cdot X + E$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:



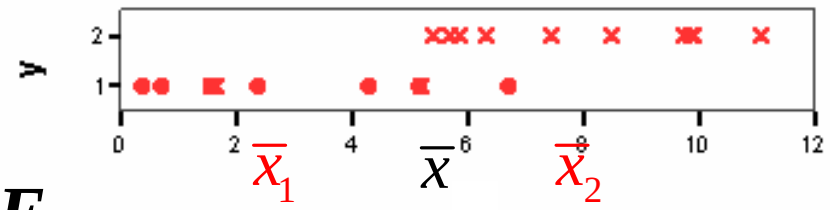
$$Y = b + a \cdot X + E$$

```
> x <- c(0.70693, 5.12851, 1.57788, 1.54994, 4.27062, 6.67777, 1.65831,
+         2.36962, 0.36968, 5.16843, 5.83867, 5.41226, 6.31614, 5.69690,
+         9.73109, 9.84856,11.06750, 7.45862, 8.47496, 9.91385)
> y <- c(rep(1,10),rep(-1,10))
> mean(x[1:10])
[1] 2.947769
> mean(x[1:10]);  mean(x[11:20]); mean(x)
[1] 2.947769
[1] 7.975855
[1] 5.461812
> var(x)*19/20
[1] 10.43800
> lm(y~x)$coefficients
(Intercept)           x
  1.3155045  -0.2408550
> round(lm(y~x)$fitted.values, digits=2)
 1.15  0.08  0.94  0.94  0.29 -0.29  0.92  0.74  1.23  0.07
-0.09  0.01 -0.21 -0.06 -1.03 -1.06 -1.35 -0.48 -0.73 -1.07
```

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:



$$Y = b + a \cdot X + E$$

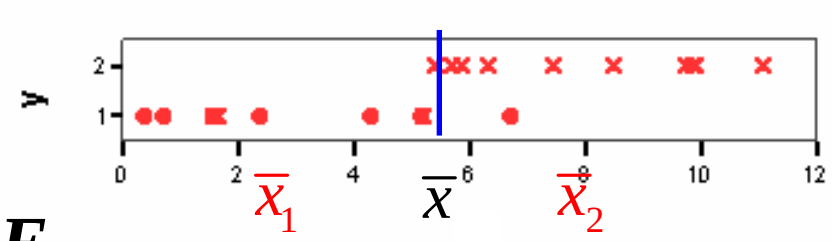
```
> x <- c(0.70693, 5.12851, 1.57788, 1.54994, 4.27062, 6.67777, 1.65831,
+         2.36962, 0.36968, 5.16843, 5.83867, 5.41226, 6.31614, 5.69690,
+         9.73109, 9.84856,11.06750, 7.45862, 8.47496, 9.91385)
> y <- c(rep(1,10),rep(-1,10))
> mean(x[1:10])
[1] 2.947769
> mean(x[1:10]);  mean(x[11:20]); mean(x)
[1] 2.947769
[1] 7.975855
[1] 5.461812
> var(x)*19/20
[1] 10.43800
> lm(y~x)$coefficients
(Intercept)           x
  1.3155045  -0.2408550
> round(lm(y~x)$fitted.values, digits=2)
 1.15  0.08  0.94  0.94  0.29 -0.29  0.92  0.74  1.23  0.07
-0.09  0.01 -0.21 -0.06 -1.03 -1.06 -1.35 -0.48 -0.73 -1.07
```

$$g(x) = 1.3155 - 0.2409 \cdot x$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:



$$Y = b + a \cdot X + E$$

```
> x <- c(0.70693, 5.12851, 1.57788, 1.54994, 4.27062, 6.67777, 1.65831,
+         2.36962, 0.36968, 5.16843, 5.83867, 5.41226, 6.31614, 5.69690,
+         9.73109, 9.84856,11.06750, 7.45862, 8.47496, 9.91385)
> y <- c(rep(1,10),rep(-1,10))
> mean(x[1:10])
[1] 2.947769
> mean(x[1:10]);  mean(x[11:20]); mean(x)
[1] 2.947769
[1] 7.975855
[1] 5.461812
> var(x)*19/20
[1] 10.43800
> lm(y~x)$coefficients
(Intercept)           x
  1.3155045  -0.2408550
> round(lm(y~x)$fitted.values, digits=2)
 1.15  0.08  0.94  0.94  0.29 -0.29  0.92  0.74  1.23  0.07
-0.09  0.01 -0.21 -0.06 -1.03 -1.06 -1.35 -0.48 -0.73 -1.07
```

$$g(x) = 1.3155 - 0.2409 \cdot x$$

Muga:

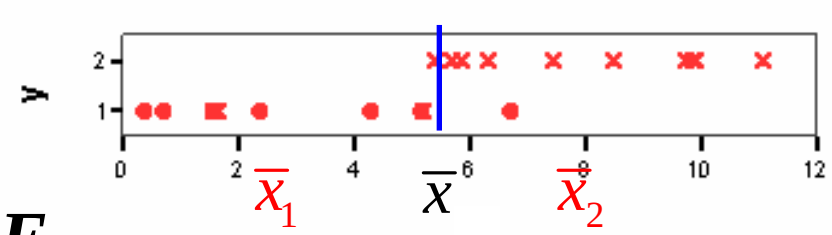
$$g(x) = 0$$

$$x_0 = 5.461812$$

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

Adibidea:



$$Y = b + a \cdot X + E$$

```
> x <- c(0.70693, 5.12851, 1.57788, 1.54994, 4.27062, 6.67777, 1.65831,
+         2.36962, 0.36968, 5.16843, 5.83867, 5.41226, 6.31614, 5.69690,
+         9.73109, 9.84856,11.06750, 7.45862, 8.47496, 9.91385)
> y <- c(rep(1,10),rep(-1,10))
> mean(x[1:10])
[1] 2.947769
> mean(x[1:10]);  mean(x[11:20]); mean(x)
[1] 2.947769
[1] 7.975855
[1] 5.461812
> var(x)*19/20
[1] 10.43800
> lm(y~x)$coefficients
(Intercept)           x
  1.3155045  -0.2408550
> round(lm(y~x)$fitted.values, digits=2)
 1.15  0.08  0.94  0.94  0.29 -0.29  0.92  0.74  1.23  0.07
-0.09  0.01 -0.21 -0.06 -1.03 -1.06 -1.35 -0.48 -0.73 -1.07
```

$$g(x) = 1.3155 - 0.2409 \cdot x$$

Muga:

$$g(x) = 0$$

$$x_0 = 5.461812$$

Orokorrean?

Analisi bereizle lineala

Linear Discriminant Analysis (LDA)

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + E$$

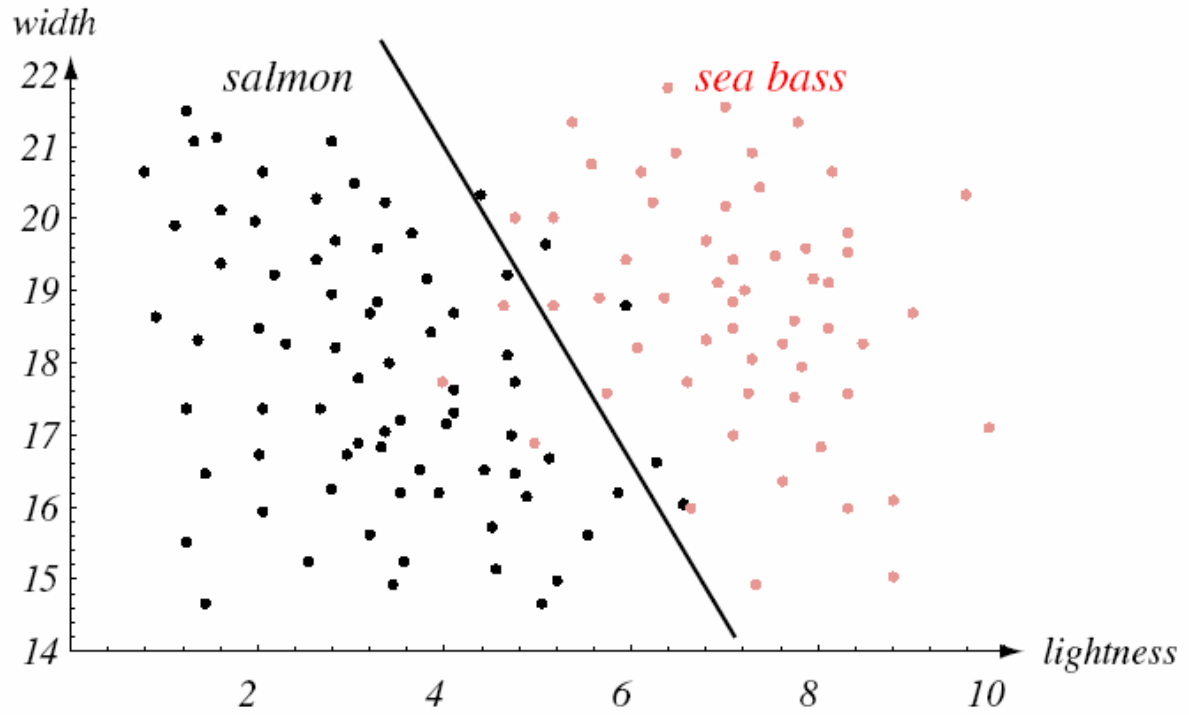


FIGURE 1.4. The two features of lightness and width for sea bass and salmon. The dark line could serve as a decision boundary of our classifier. Overall classification error on the data shown is lower than if we use only one feature as in Fig. 1.3, but there will still be some errors. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.