Confusion matrix

TP FP FN TN

		Actual	
		Cancer (P)	Non-cancer (N)
Predicted	Cancer	45	7
	Non-cancer	2	40

$$\text{Accuracy} = \frac{\#\text{ correct}}{\#\text{ total}} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{85}{94}$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{45}{52}$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{45}{47}$$

Actual

Predicted	Hares	Bears	Wolves
Hares	21	6	3
Bears	3	15	8
Wolves	5	1	25

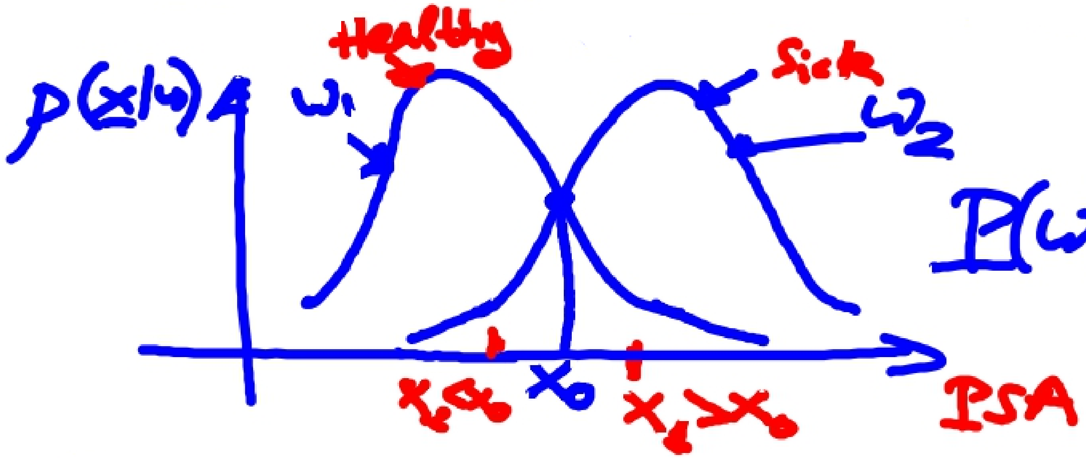
$$\text{Accuracy} = \frac{21+15+25}{21+6+3+3+15+8+5+1+25} = \frac{61}{87}$$

$$\text{Precision (Hares)} = \frac{21}{30}$$

$$\text{Recall(Hares)} = \frac{21}{29}$$

Bayes classifier

Classify to $\omega_{i(x)}$: $\quad P(\omega_1|\underline{x}) \gtrless P(\omega_2|\underline{x})$



$$P(\omega_1)=P(\omega_2).$$

Risk: Substitute P_e taking into account risk.

Risk to have decided that $\underline{x}$ is in ω_1, whereas the truth is ω_2

$$r_2 = \lambda_{21} \int_{R_1} p(\underline{x}|\omega_2)\, d\underline{x}$$

$$r_1 = \lambda_{12} \int_{R_2} p(\underline{x}|\omega_1)\, d\underline{x}$$

$$r = P(\omega_1)\, r_1 + P(\omega_2)\, r_2$$

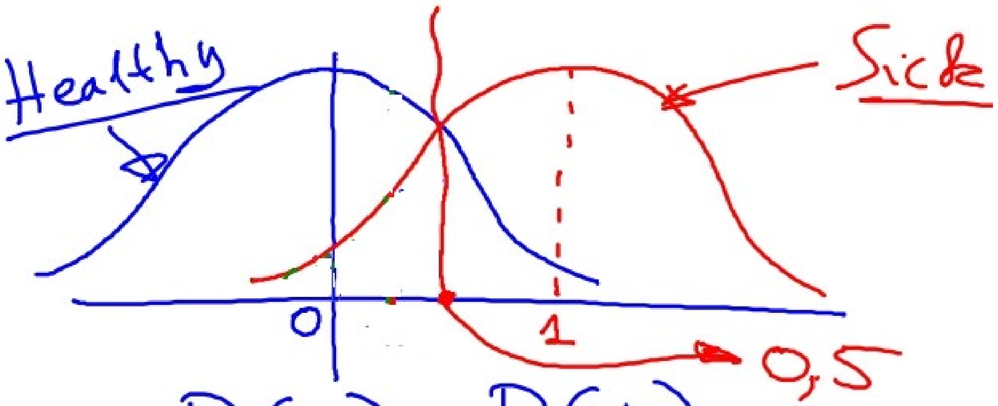
$$r = \lambda_{12} P(\omega_1) - \int_{R_1} \left[\lambda_{12} P(\omega_1|\underline{x}) - \lambda_{21} P(\omega_2|\underline{x}) \right] p(\underline{x})\, d\underline{x}$$

Classify in ω_1: $\lambda_{12} P(\omega_1|\underline{x}) > \lambda_{21} P(\omega_2|\underline{x})$

$$\boxed{\frac{p(\underline{x}|\omega_1)}{p(\underline{x}|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \frac{\lambda_{21}}{\lambda_{12}}}$$

Example:

$$p(x|\omega_1) = \frac{1}{\sqrt{\pi}} e^{-x^2/2} \qquad p(x|\omega_2) = \frac{1}{\sqrt{\pi}} e^{-\frac{(x-1)^2}{2}}$$



$$P(\omega_1) = P(\omega_2).$$

$$\lambda_{12} = 0,5$$

$$\lambda_{21} = 1.$$

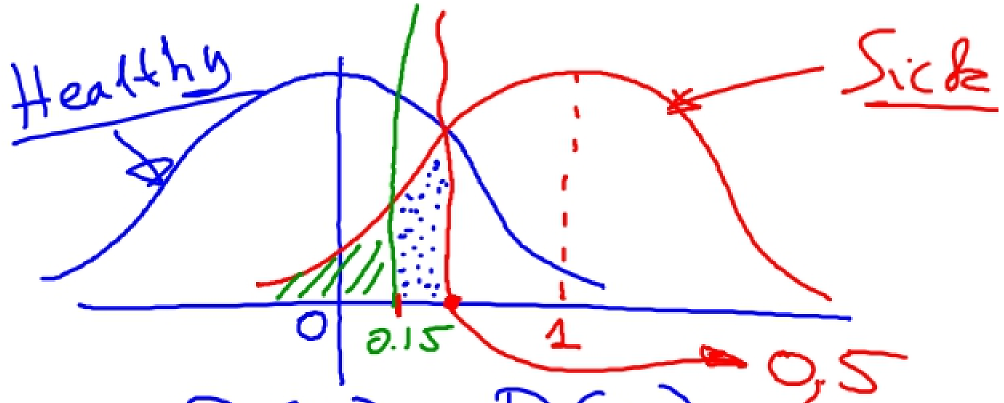
$\longrightarrow$ Importance of having chosen ω_1, whereas truth is ω_2.

Separating point:

$$\frac{e^{-x^2}}{e^{-(x-1)^2}} = \frac{1}{0,5} = 2 \Rightarrow e^{-x^2} = 2e^{-(x-1)^2}$$

$$\Rightarrow -x^2 = \ln 2 - (x-1)^2 \Rightarrow \cancel{-x^2} = \ln 2 - \cancel{x^2} - 1 + 2x$$

$$\Rightarrow x = \frac{-\ln 2 + 1}{2} \simeq 0,15$$



$$P(\omega_1) = P(\omega_2).$$

Bayes classifier: The Gaussian case.

$$p(\underline{x}|\omega_i) = \frac{1}{(2\pi)^{\ell/2}(\det \Sigma_i)^{-1/2}} \cdot \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu}_i)^T \Sigma_i^{-1}(\underline{x}-\underline{\mu}_i)\right]$$

$$p(\underline{x}|\omega_i)P(\omega_i) \qquad i = 1, 2.$$

Discriminant function:

$$g_i(\underline{x}) = \ln\left[p(\underline{x}|\omega_i)P(\omega_i)\right].$$

Decision surface:

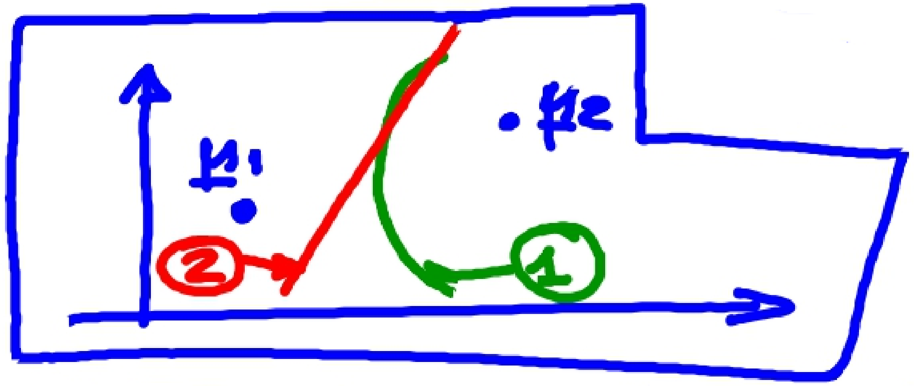
$$g_1(\underline{x}) - g_2(\underline{x}) = 0$$

$$-\underline{x}^T \Sigma_1^{-1} \underline{x} + \underline{x}^T \Sigma_1^{-1} \underline{\mu}_1 + \underline{\mu}_1 \Sigma_1^{-1} \underline{x} - \underline{\mu}_1 \Sigma_1^{-1} \underline{\mu}_1 + 2\ln P(\omega_1) - \ln \det \Sigma_1$$

$$= -\underline{x}^T \Sigma_2^{-1} \underline{x} + \underline{x}^T \Sigma_2^{-1} \underline{\mu}_2 + \mu_2^T \Sigma_2^{-1} \underline{x} - \underline{\mu}_2^T \Sigma_2^{-1} \underline{\mu}_2 + 2\ln P(\omega_2) - \ln \det \Sigma_2$$

1. $\Sigma_1 \neq \Sigma_2 \Rightarrow$ Quadratic decision surface (ellipsoid, paraboloid...)

2. $\Sigma_1 = \Sigma_2 \Rightarrow$ hyperplane.



2a $P(\omega_1)=P(\omega_2)$

$\Sigma=\Sigma_1=\Sigma_2$ $\Rightarrow$ $\underline{x}$ in ω_1 iff:

$$d_1^2=(\underline{x}-\underline{\mu}_1)^T\Sigma^{-1}(\underline{x}-\underline{\mu}_1)$$
$$<(\underline{x}-\underline{\mu}_2)^T\Sigma^{-1}(\underline{x}-\underline{\mu}_2)=d_2^2$$

d_1, d_2 : <u>Mahalanobis distance.</u>

2b $\Sigma_1=\Sigma_2=\sigma^2\mathbb{I}$ $\Rightarrow$ <u>Euclidian distance.</u>

More on geometry

Ⓐ $\Sigma_1 = \Sigma_2 = \sigma^2 \mathbb{I}$

Decision surface:

$$\ln P(\omega_1) - \frac{1}{2}\frac{\|\underline{x}-\underline{\mu}_1\|^2}{\sigma^2} = \ln P(\omega_2) - \frac{1}{2}\frac{\|\underline{x}-\underline{\mu}_2\|^2}{\sigma^2}$$

$$\Rightarrow \sigma^2 \ln\frac{P(\omega_1)}{P(\omega_2)} - \frac{1}{2}\Big(\cancel{\|\underline{x}\|^2} - 2\underline{\mu}_1^T\underline{x} + \|\underline{\mu}_1\|^2 - \cancel{\|\underline{x}\|^2} + 2\underline{\mu}_2^T\underline{x} - \|\underline{\mu}_2\|^2\Big) = 0$$

$$\Rightarrow (\underline{\mu}_1 - \underline{\mu}_2)^T\underline{x} - \frac{1}{2}\left(\|\underline{\mu}_1\|^2 - \|\underline{\mu}_2\|^2\right) + \sigma^2 \ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right] = 0$$

$$\Rightarrow \quad (\underline{\mu}_1-\underline{\mu}_2)^T \underline{x} - \frac{1}{2}\left(\|\underline{\mu}_1\|^2 - \|\underline{\mu}_2\|^2\right) + \sigma^2 \ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right] = 0$$

Want to put it in the form: $\underline{w}^T(\underline{x}-\underline{x}_0) = 0$

$\underline{w} = \underline{\mu}_1 - \underline{\mu}_2$

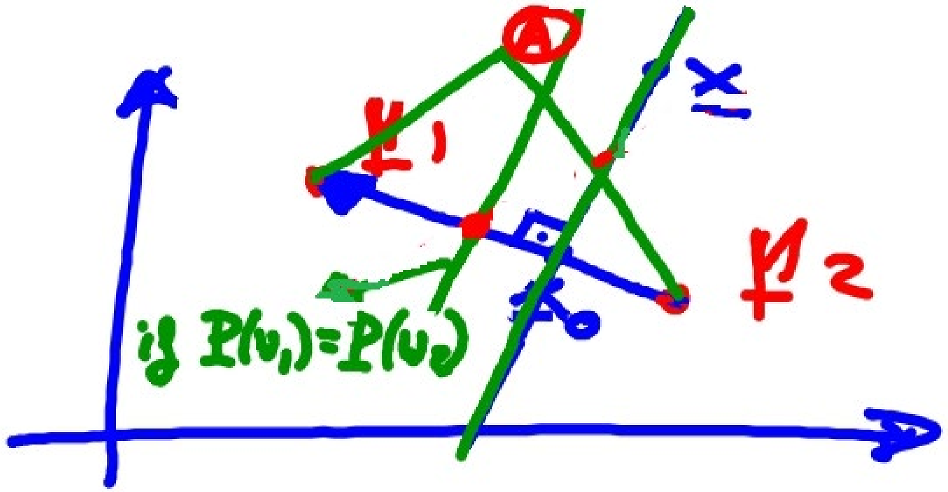
$$\underline{x}_0 = \frac{1}{2}(\underline{\mu}_1+\underline{\mu}_2) - \sigma^2 \ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right] \frac{\underline{\mu}_1-\underline{\mu}_2}{\|\underline{\mu}_1-\underline{\mu}_2\|^2}$$

$$\underline{w}^T \underline{x}_0 = \underbrace{\frac{1}{2}(\underline{\mu}_1-\underline{\mu}_2)^T(\underline{\mu}_1+\underline{\mu}_2)}_{\|\underline{\mu}_1\|^2 - \|\underline{\mu}_2\|^2} - \sigma^2 \ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right] \frac{\cancel{(\underline{\mu}_1-\underline{\mu}_2)^T(\underline{\mu}_1-\underline{\mu}_2)}}{\cancel{\|\underline{\mu}_1-\underline{\mu}_2\|^2}}$$

$$\boxed{(\underline{\mu}_1-\underline{\mu}_2)^T(\underline{x}-\underline{x}_0)=0.}$$

$$\underline{x}_0=\frac{1}{2}(\underline{\mu}_1+\underline{\mu}_2)-\sigma^2\ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right]\frac{\underline{\mu}_1-\underline{\mu}_2}{\|\underline{\mu}_1-\underline{\mu}_2\|^2}$$

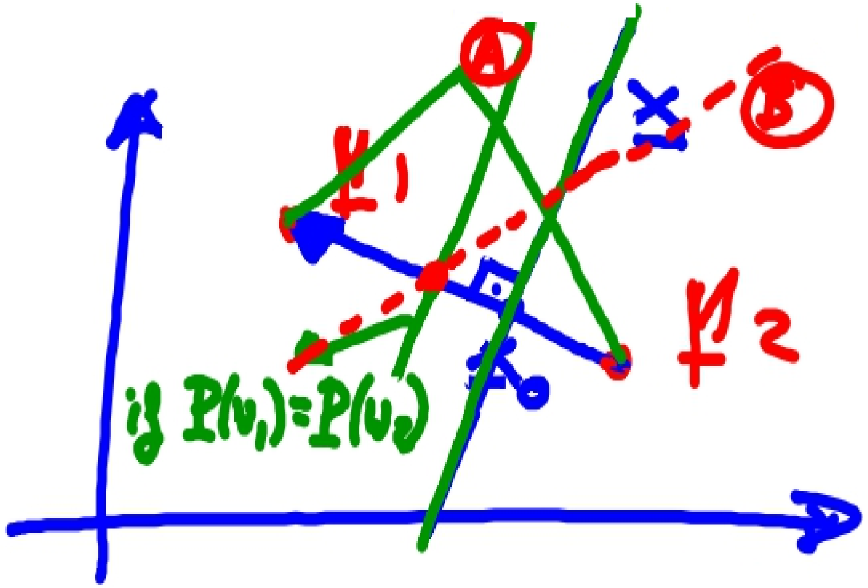
$P(\omega_1)=P(\omega_2)$: Separating hyperplane: Perpendicular bisector of $\underline{\mu}_1-\underline{\mu}_2$.



$$\text{B} \quad \Sigma_1 = \Sigma_2 \neq \sigma^2 \mathbb{I}$$

Separating hyperplane:

$$\underline{w}^T(\underline{x} - \underline{x}_0) = 0. \qquad \underline{w} = \Sigma^{-1}(\mu_1 - \mu_2)$$



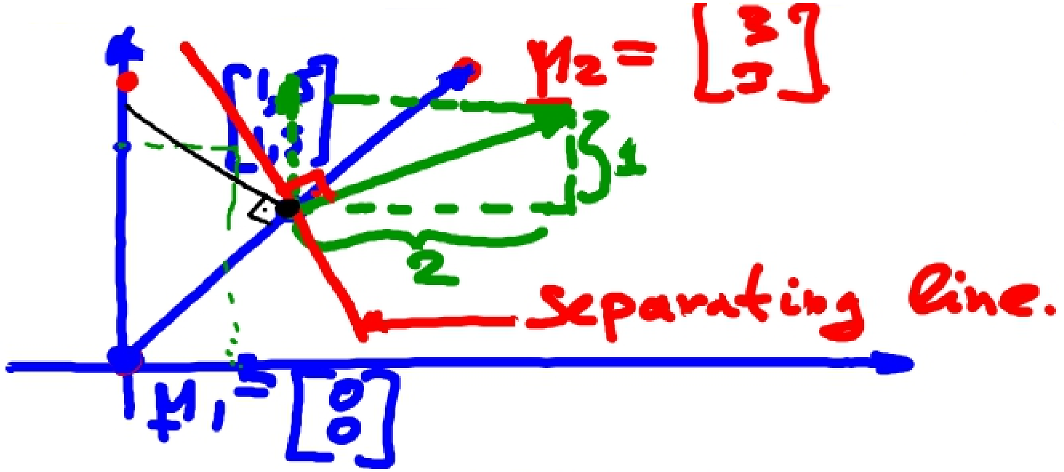
$$x_0 = \frac{1}{2}(\mu_1 + \mu_2) - \ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right] \cdot \frac{\mu_1 - \mu_2}{\|\mu_1 - \mu_2\|^2_{\Sigma^{-1}}}$$

$$\|y\|_{\Sigma^{-1}} = (y^T \Sigma^{-1} y)^{1/2}$$

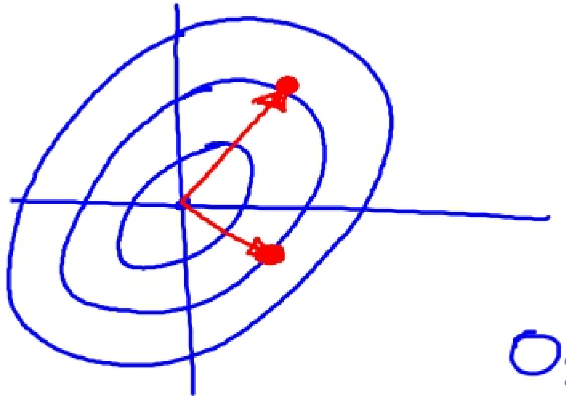
Example (continued):

$$\Sigma = \begin{bmatrix} 1,1 & 0,3 \\ 0,3 & 1,9 \end{bmatrix}$$

$$\Rightarrow \Sigma^{-1} = \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix}$$



$$\underline{w} = \Sigma^{-1}(\mu_1 - \mu_2) = \begin{bmatrix} 0,95 & -0,15 \\ -0,15 & 0,55 \end{bmatrix} \begin{bmatrix} -3 \\ -3 \end{bmatrix} = 1,2 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$



$$\underline{z}^T \Sigma^{-1} \underline{z} = c$$

difference from the mean.

Optimize $\underline{z}^T \underline{z}$ s.t. $\underline{z}^T \Sigma^{-1} \underline{z} = c$

$$\mathcal{L} = \underline{z}^T \underline{z} - \lambda \left(\underline{z}^T \Sigma^{-1} \underline{z} - c \right) \qquad \frac{d\mathcal{L}}{d\underline{z}} = 0 \Rightarrow \underline{z} = \lambda \Sigma^{-1} \underline{z}$$

$$\Rightarrow \Sigma \underline{z} = \lambda \cancel{\Sigma} \Sigma^{-1} \underline{z} \Rightarrow \Sigma \underline{z} = \lambda \underline{z},$$ $\underline{z}$ is eigenvector of Σ.

$$\underline{z}^T \Sigma^{-1} \underline{z} = c \Rightarrow \underline{z}^T \underline{z} \frac{1}{\lambda} = c \Rightarrow \|\underline{z}\| = \sqrt{c}\sqrt{\lambda}$$

Example again:

$$\Sigma = \begin{bmatrix} 1,1 & 0,3 \\ 0,3 & 1,9 \end{bmatrix}.$$

Eigenvalues: $\det(\Sigma - \lambda \mathbb{I}) = 0$

$$(1,1-\lambda)(1,9-\lambda) - 0,09 = 0$$

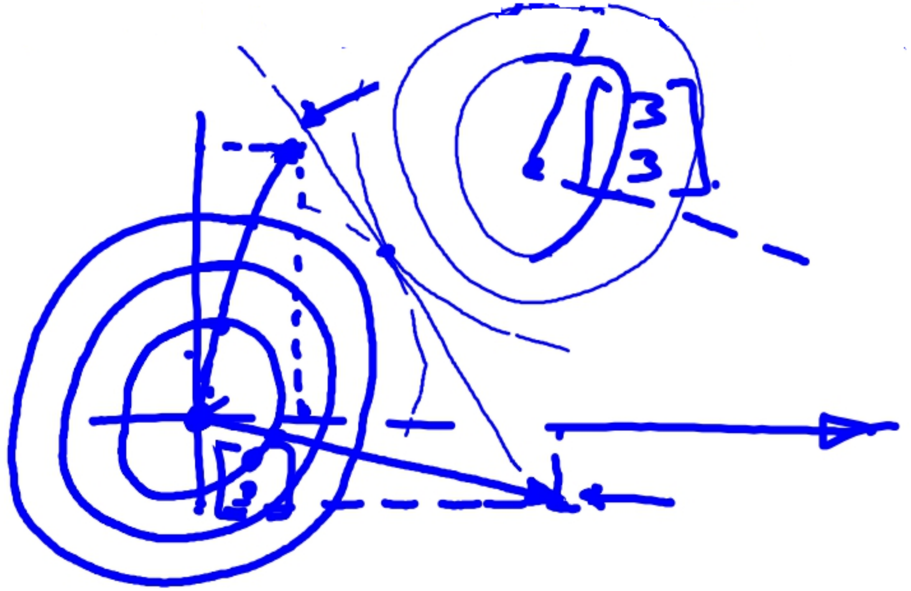
$$\lambda = \begin{cases} \lambda_1 = 2 \\ \lambda_2 = 1. \end{cases}$$

$\lambda_1 = 2$: Eigenvector: $\begin{bmatrix} 1{,}1 & 0{,}3 \\ 0{,}3 & 1{,}9 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 2 \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$

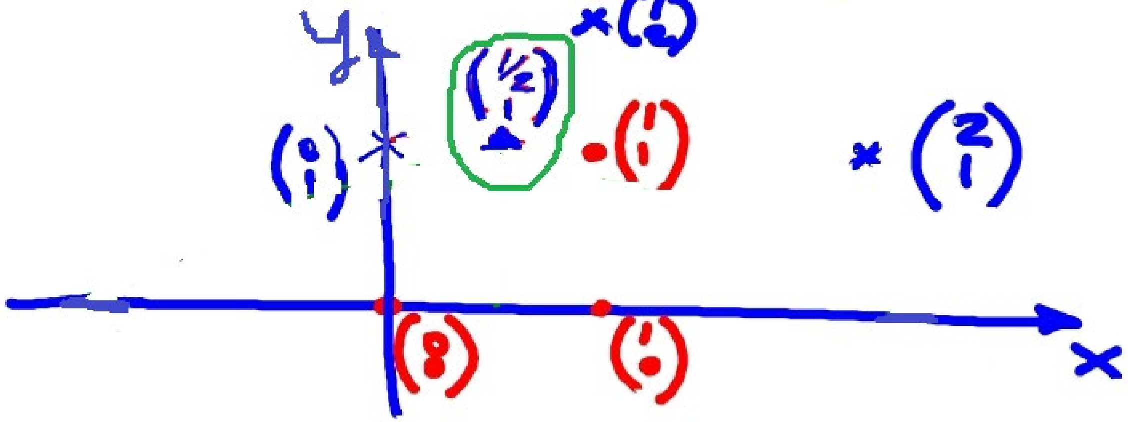
$\Rightarrow 1{,}1 z_1 + 0{,}3 z_2 = 2 z_1 \Rightarrow z_2 = 3 z_1 \Rightarrow \underline{z} = \begin{bmatrix} k \\ 3k \end{bmatrix}$

Normalize: $\|\underline{z}\|^2 = 1 \Rightarrow k^2 + 3^2 k^2 = 1 \Rightarrow k = \frac{1}{\sqrt{10}} \Rightarrow \underline{z} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

$\lambda_2 = 1$: Eigenvector: $\underline{z} = \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \end{bmatrix}$.



Example: Naïve Bayes classifier.



A: Red class

B: Blue class

A:

x: 0 1 1

y: 0 0 1

$E(x) = \frac{2}{3}, \; E(y) = \frac{1}{3}.$

$\sigma^2(x) = \frac{1}{3}\left[\left(0-\frac{2}{3}\right)^2 + 2\left(1-\frac{2}{3}\right)^2\right]$

$= 2/9$

$\sigma^2(y) = 2/9$

B:

x: 0 1 2

y: 1 2 1

$E(x) = 1, \; E(y) = \frac{4}{3}$

$\sigma^2(x) = 2/3$

$\sigma^2(y) = 2/9.$

$$P_A(x,y) = \frac{1}{2\pi\sqrt{2/9}\sqrt{2/9}}\, e^{-\frac{\left(x-\frac{2}{3}\right)^2}{2\cdot 2/9}}\, e^{-\frac{\left(y-\frac{1}{3}\right)^2}{2\cdot 2/9}}$$

$$= \frac{9}{4\pi} \exp\left[-\frac{9}{4}\left(x-\frac{2}{3}\right)^2 - \frac{9}{4}\left(y-\frac{1}{3}\right)^2\right]$$

$$P_B(x,y) = \frac{3\sqrt{3}}{4\pi} \exp\left[-\frac{3(x-1)^2}{4} - \frac{9\left(y-\frac{4}{3}\right)^2}{4}\right].$$

<u>Decision curve:</u>

$$\ln P_A(x,y) = \ln P_B(x,y)$$

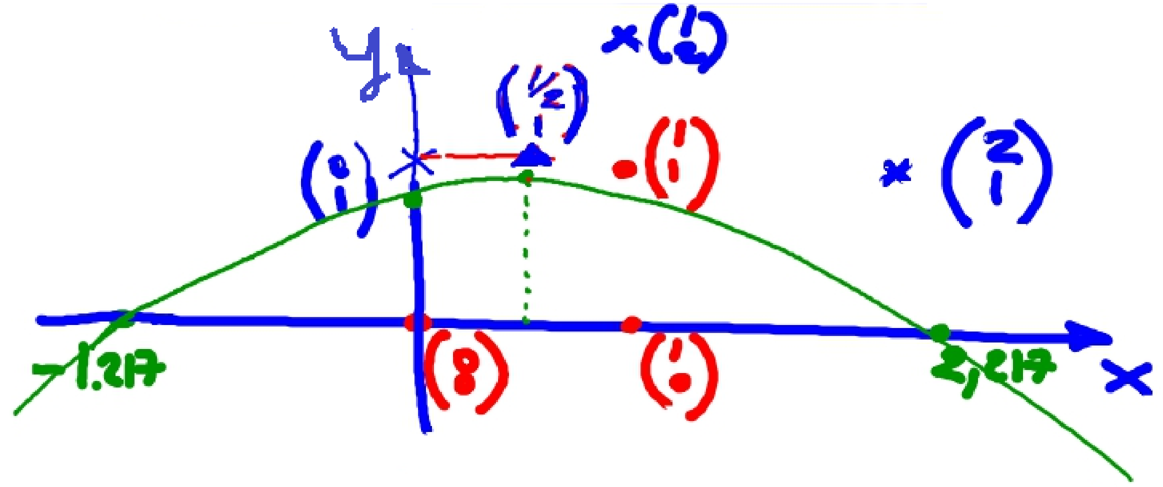
$$\Rightarrow \ln 3 = 3x^2 - 3x + 9y - 7$$

$$\Rightarrow \quad y = \frac{7+\ln 3}{9} + \frac{x-x^2}{3}.$$

$$x=0 \Rightarrow y = \frac{7+\ln 3}{9} \simeq 0{,}9$$

$$y=0 \Rightarrow -x^2+x+\frac{7+\ln 3}{3}=0 \Rightarrow \begin{cases} x_1 = -1.217 \\ x_2 = 2.217 \end{cases}$$

<u>Maximum:</u> $\frac{dy}{dx}=0 \Rightarrow \frac{1}{3}-\frac{2x}{3}=0 \Rightarrow x=\frac{1}{2} \Rightarrow y=0{,}983.$



Classify $\begin{pmatrix} 1/2 \\ 1 \end{pmatrix}$

$$\ln P_A\left(\tfrac{1}{2}, 1\right) = 1{,}1347$$

$$\ln P_B\left(\tfrac{1}{2}, 1\right) = 1{,}2104.$$

$\Rightarrow \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}$ goes to class B.