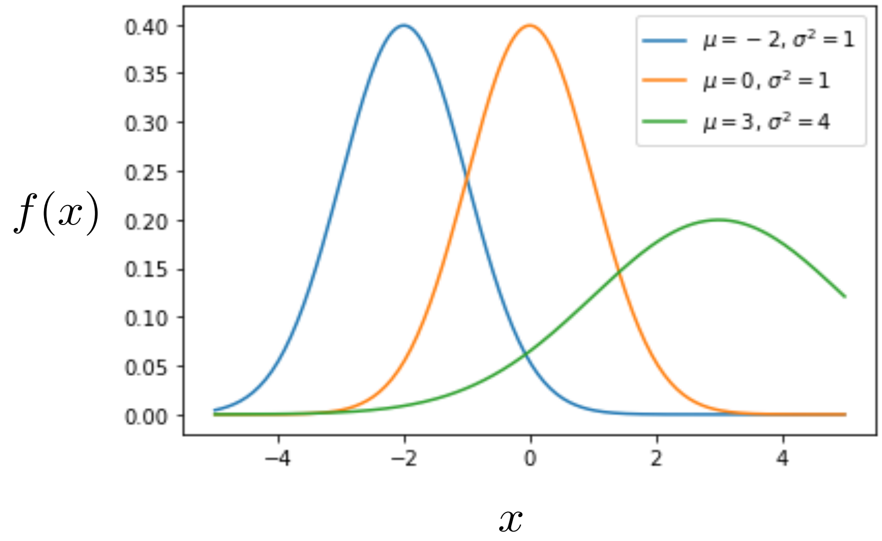
Gaussian Random Vectors, Gaussian Discriminant Analysis

- multivariate normal distribution
- Gaussian discriminant analysis

Gaussian Distribution

- Gaussian: $X \sim \mathcal{N}(\mu, \sigma^2)$

$$f(x) = \frac{\exp\left[\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}{\sqrt{2\pi}\,\sigma}$$



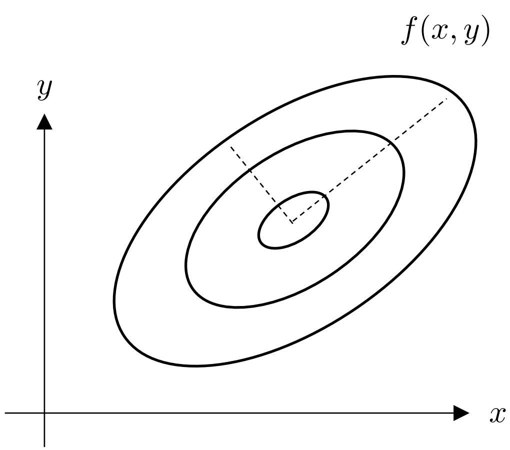
- Bivariate (jointly) Gaussian:

$$f(x,y) = \frac{\exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}\right]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

$\mu_x := E[X] \qquad \sigma_x^2 := \mathrm{var}(X)$

$\mu_y := E[Y] \qquad \sigma_y^2 := \mathrm{var}(Y)$

$\rho :=$ correlation



Multivariate Normal

- a random vector $\boldsymbol{x} \in \mathbb{R}^n$ with mean $\boldsymbol{\mu}$ and covariance matrix Σ is a (jointly) Gaussian random vector if and only if its pdf is given by:

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)$$

where $|\boldsymbol{\Sigma}|$ is the determinant of Σ.

- $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates $\boldsymbol{x}$ is a Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

$$E[\boldsymbol{x}] = \boldsymbol{\mu} \qquad \mathrm{cov}(\boldsymbol{x}) = E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T] = \boldsymbol{\Sigma}$$

- Σ^{-1} is positive definite because Σ is positive definite; i.e.,

$$(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \geq 0, \qquad \text{with strict inequality for } \boldsymbol{x} \neq \boldsymbol{\mu}.$$

- contours are given by ellipses

$$(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) = c$$

Multivariate Normal

- contours are given by ellipses

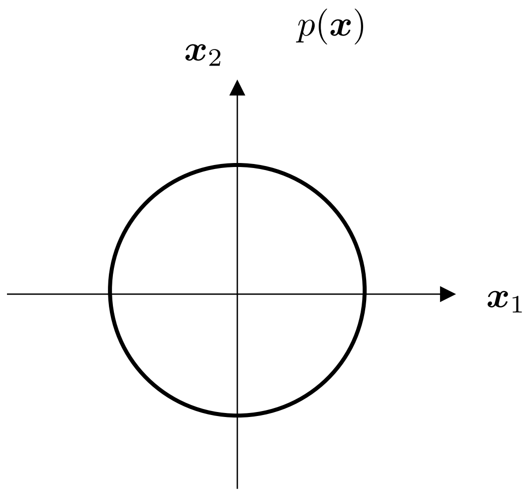
$$(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c$$

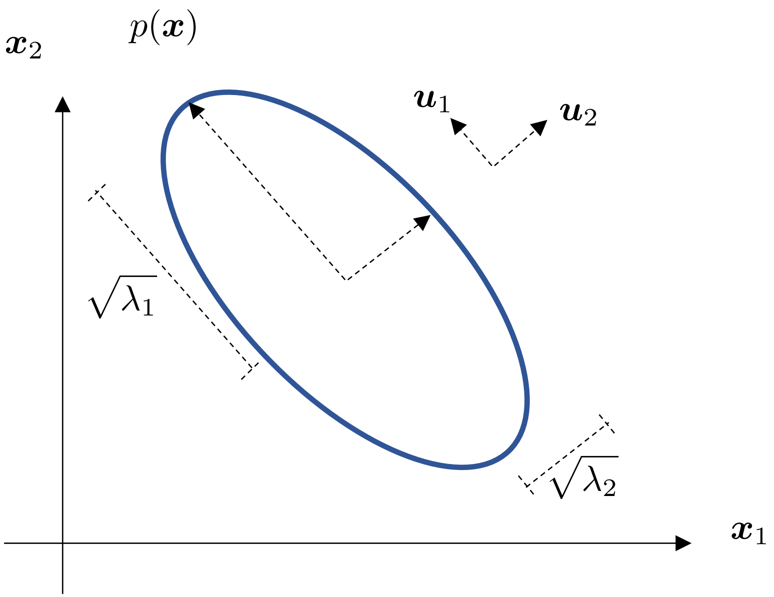
$$\boldsymbol{\mu} = 0, \boldsymbol{\Sigma} = \boldsymbol{I}$$

$$\boldsymbol{x}^T\boldsymbol{x} = ||\boldsymbol{x}||_2^2 = c$$

$$\boldsymbol{\Sigma} = \boldsymbol{U}\Lambda\boldsymbol{U}^T = \sum \lambda_i \boldsymbol{u}\boldsymbol{u}^T$$

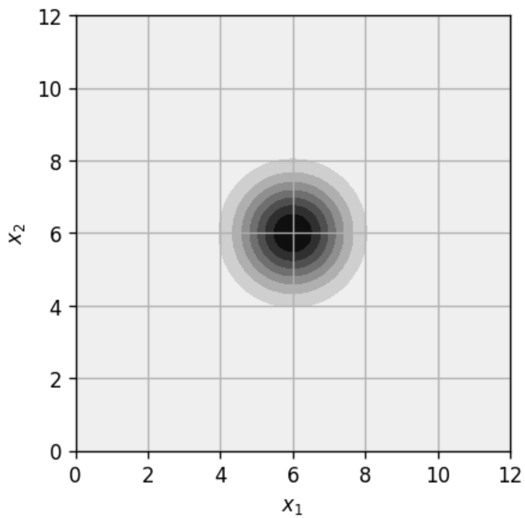
$$(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \sum \frac{1}{\lambda_i}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{u}_i\boldsymbol{u}_i^T(\boldsymbol{x}-\boldsymbol{\mu})$$
$$= \sum \frac{y_i^2}{\lambda_i}$$

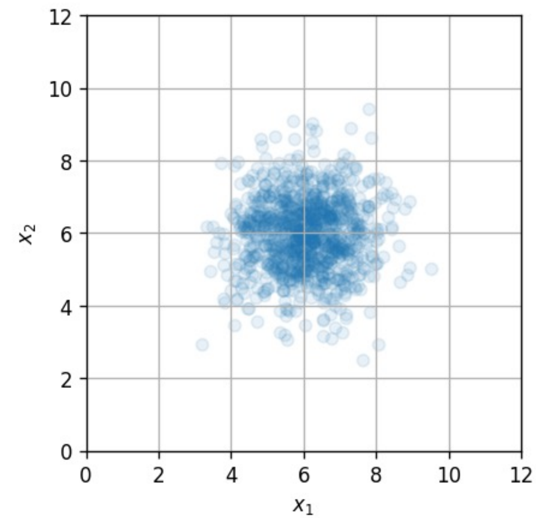




Multivariate Normal

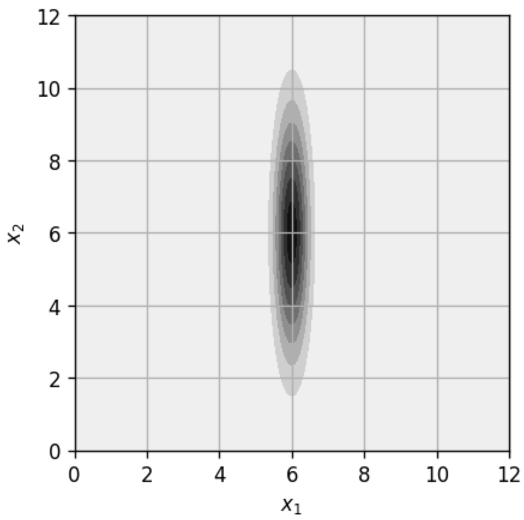
$$\boldsymbol{\mu} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

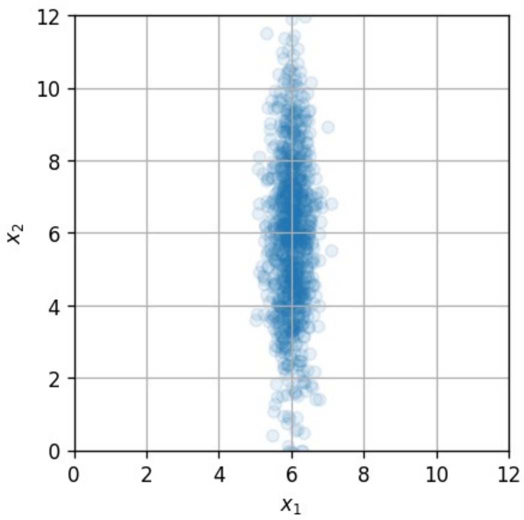




Multivariate Normal

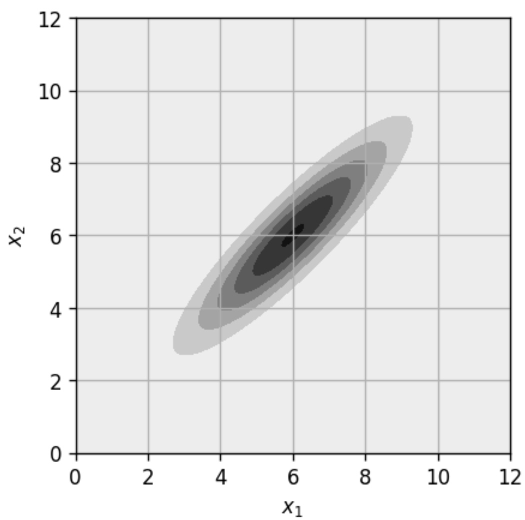
$$\boldsymbol{\mu} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 0.1 & 0 \\ 0 & 5 \end{bmatrix}$$

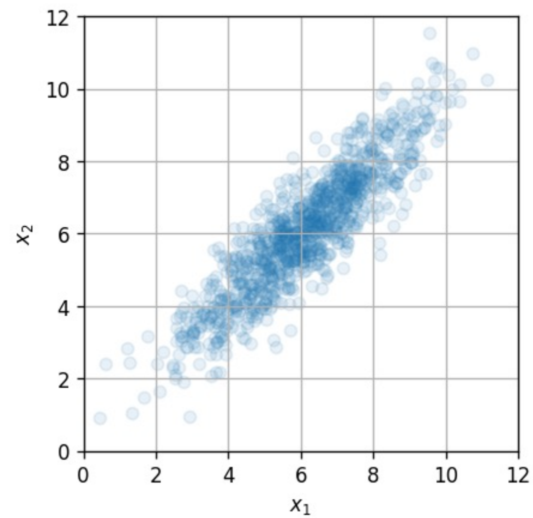




Multivariate Normal

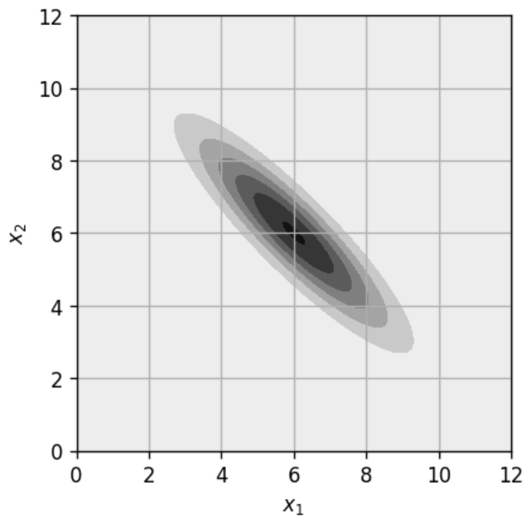
$$\boldsymbol{\mu} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 3 & 2.7 \\ 2.7 & 3 \end{bmatrix}$$

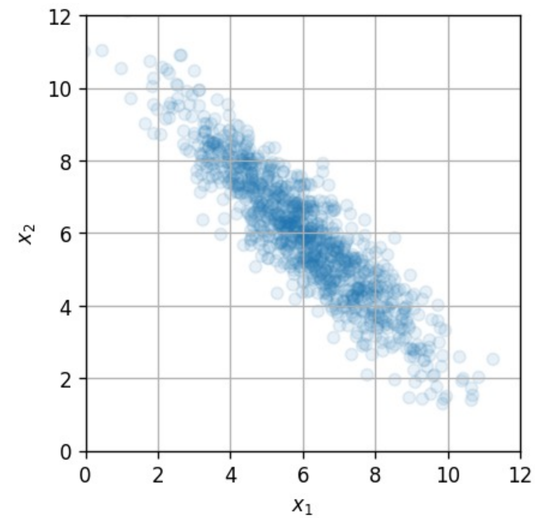




Multivariate Normal

$$\boldsymbol{\mu} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 3 & -2.7 \\ -2.7 & 3 \end{bmatrix}$$





Multivariate Normal Properties

• property 1: linear transformation of Gaussian random vectors are Gaussian random vectors:

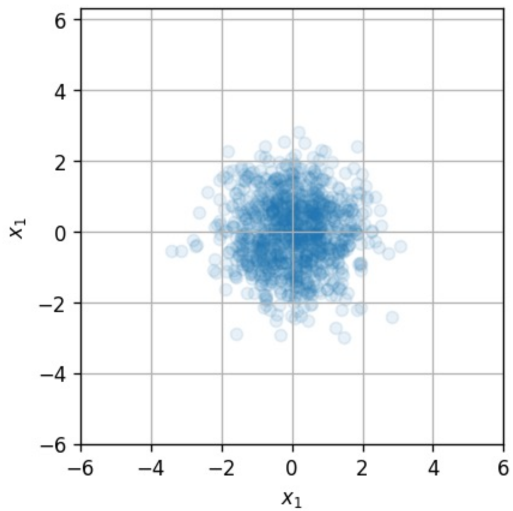
$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{A}\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T)$$

• we also need to show that $\boldsymbol{A}\boldsymbol{x}$ is a Gaussian random vector (proof in note)

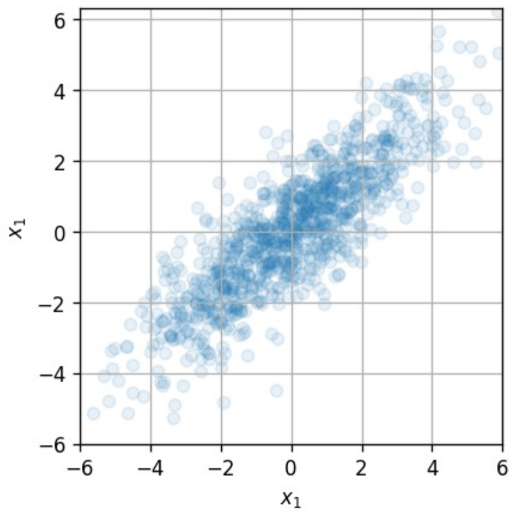
Multivariate Normal

$$\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$$



$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix}2 & 0\\1.5 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

$$\mathcal{N}\left(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T\right)$$



Multivariate Normal Properties

- property 2: for Gaussian random vectors, uncorrelated $\Longleftrightarrow$ independence

  uncorrelated: $\boldsymbol{\Sigma} = \boldsymbol{I}$

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\sum_i (x_i - \mu_i)^2\right)$$

$$= \prod_i \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_i - \mu_i)^2\right)$$

Multivariate Normal Properties

- property 3: marginals are also Gaussian random vectors

$$\boldsymbol{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

- follows from property 1 by taking $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ with appropriate $\boldsymbol{A}$.

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \qquad \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{A}\mu, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T)$$

- example - $p(x_1, x_2)$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \end{bmatrix}$$

- we just remove marginalized indexes from the mean and covariance matrix

Multivariate Normal Properties

- property 4: conditionals are also Gaussian random vectors

$$\boldsymbol{x} = \left[\begin{array}{c} \boldsymbol{x}_1 \\ -- \\ \boldsymbol{x}_2 \end{array}\right] \qquad \boldsymbol{x} \sim \mathcal{N}\left(\left[\begin{array}{c} \boldsymbol{\mu}_1 \\ -- \\ \boldsymbol{\mu}_2 \end{array}\right], \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ -- & -- \\ \Sigma_{21} & \Sigma_{22} \end{array}\right]\right)$$

then

$$\boldsymbol{x}_2|\boldsymbol{x}_1 \sim \mathcal{N}\left(\Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu_1}) + \boldsymbol{\mu_2}, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$$

$\Sigma_{x_2|x_1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ is the *Schur* complement of Σ_{11}

- *Schur complement* invert Σ, drop rows and columns that are conditioned up, and then invert smaller matrix

- we won't prove this – see, i.e, [KPM 4.3.4.3]

Example –

- example:

$$\boldsymbol{x} \sim \mathcal{N}\left(\begin{bmatrix}3\\0\\1\end{bmatrix}, \begin{bmatrix}2 & 1 & 2\\1 & 7 & 2\\2 & 2 & 3\end{bmatrix}\right)$$

- what is $p(x_2)$?

$$p(x_2) = \int_{-\infty}^{-\infty}\int_{-\infty}^{-\infty} p(x_1, x_2, x_3) dx_1 dx_3$$

$$p(x_2) = \int_{-\infty}^{-\infty}\int_{-\infty}^{-\infty} \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mu)\right) dx_1 dx_3$$

- a better approach:

set $\boldsymbol{A} = [0\ 1\ 0]$

$$\boldsymbol{A}\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{A}\mu, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T)$$

$$p(x_2) \sim \mathcal{N}(0, 7)$$

Example –

- example:

$$\boldsymbol{x} \sim \mathcal{N}\left(\begin{bmatrix}3\\0\\1\end{bmatrix}, \begin{bmatrix}2 & 1 & 2\\1 & 7 & 2\\2 & 2 & 3\end{bmatrix}\right) \qquad \boldsymbol{x} = \left[\begin{array}{c}\boldsymbol{x}_1\\ --\\ \boldsymbol{x}_2\end{array}\right] \qquad \mathcal{N}\left(\left[\begin{array}{c}\boldsymbol{\mu}_1\\ --\\ \boldsymbol{\mu}_2\end{array}\right], \left[\begin{array}{c|c}\Sigma_{11} & \Sigma_{12}\\ -- & --\\ \Sigma_{21} & \Sigma_{22}\end{array}\right]\right)$$

$$\boldsymbol{x}_2|\boldsymbol{x}_1 \sim \mathcal{N}\left(\Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu_1}) + \boldsymbol{\mu_2}, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$$

- what is $p(x_2, x_3|x_1)$?

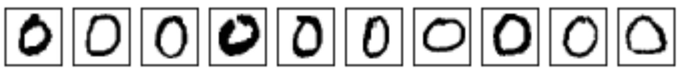
Gaussian Discriminant Analysis

$$\widehat{y} = \arg\max_{y} p(\boldsymbol{x}|y)$$

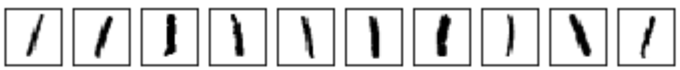
MAP

ML (maximum likelihood) estimate

- generative classifier: assume that x was generated by a class conditional distribution:



$$p(\boldsymbol{x}|y=0) \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$



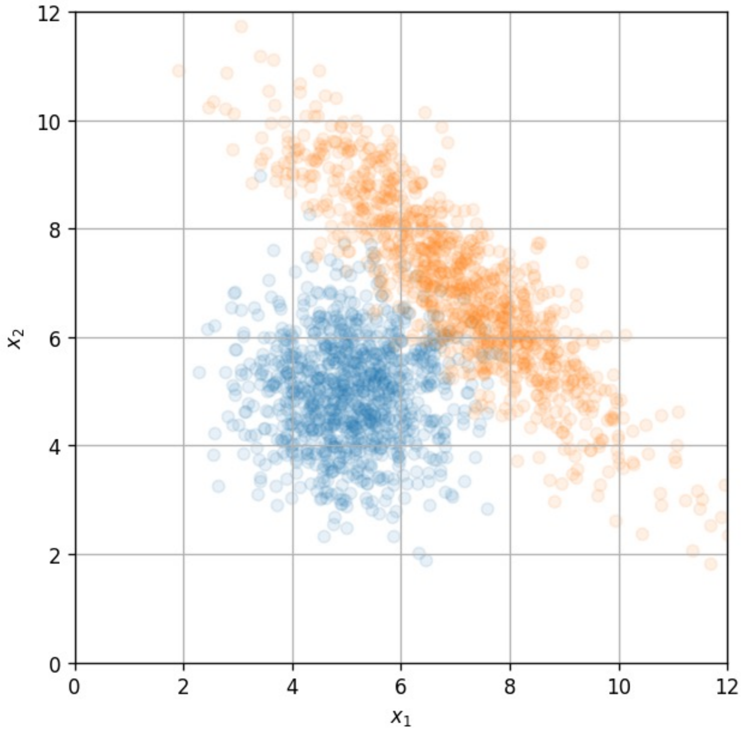
$$p(\boldsymbol{x}|y=1) \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$$

- given a new $\boldsymbol{x}$, we can evaluate $p(\boldsymbol{x}|y=0)$ vs. $p(\boldsymbol{x}|y=1)$

Gaussian Discriminant Analysis

$$\boldsymbol{x}|y=0 \sim \mathcal{N}\left(\begin{bmatrix}5\\5\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right)$$

$$\boldsymbol{x}|y=1 \sim \mathcal{N}\left(\begin{bmatrix}7\\7\end{bmatrix}, \begin{bmatrix}3 & -2.7\\-2.7 & 3\end{bmatrix}\right)$$

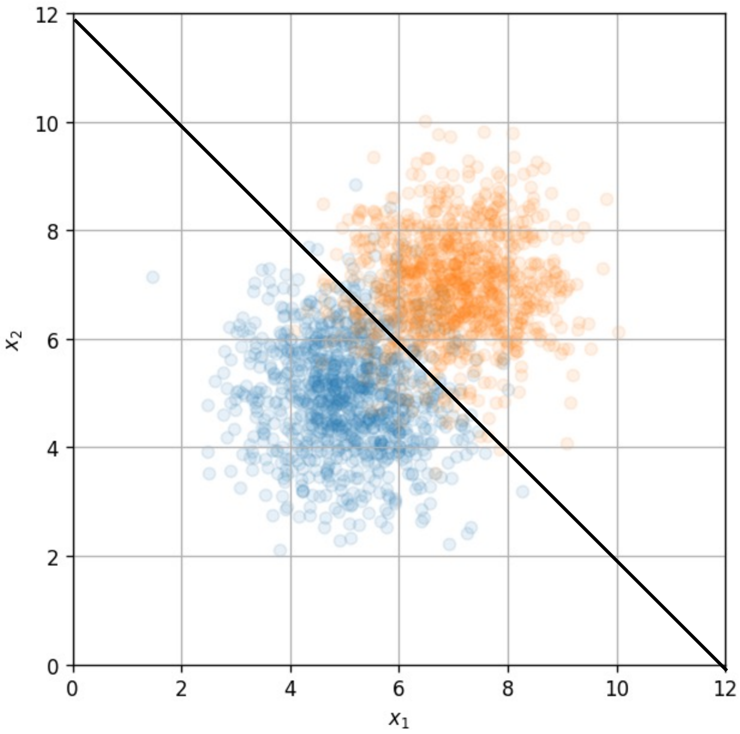


$$\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\right) \quad \text{vs.} \quad \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\right)$$

Linear Discriminant Analysis

- linear discriminant analysis: $\Sigma_0 = \Sigma_1$

$$\frac{\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\right)}{\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\right)} \gtrless 1$$



- decision boundary has a linear form:

$$\boldsymbol{x}^T\boldsymbol{w} \gtrless c$$

Gaussian Discriminant Analysis

- quadratic discriminant analysis: $\Sigma_0 \neq \Sigma_1$

$$\frac{\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\right)}{\frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\right)} \lesseqgtr 1$$

- decision boundary has a quadratic form:

$$\boldsymbol{x}^T B\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{w} \lesseqgtr c$$