# Gaussian Random Vectors, Gaussian Discriminant Analysis

*Submit a PDF of your answers to Canvas.*

**1.** If a random vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a Gaussian random vector, then the marginals are Gaussian, i.e, $p(x_j) \sim \mathcal{N}(\mu_j, \sigma_j)$. In this problem, we consider the converse. Let $X_1 \sim \mathcal{N}(0, 1)$ and define the random variable

$$X_2 = \begin{cases} X_1 & \text{with probability } 1/2 \\ -X_1 & \text{with probability } 1/2. \end{cases}$$

**a)** What is $p(x_2)$?

**b)** Sketch $p(x_1, x_2)$ in the $x_1, x_2$ plane (or describe the joint distribution of $X_1, X_2$).

**c)** Is $[X_1\ X_2]^T$ a Gaussian random vector?

**SOLUTION:**

**a)** $p(x_2) \sim \mathcal{N}(0, 1)$.

**b)** By construction, $x_2 = -x_1$ or $x_2 = x_1$, so the pdf has all its mass on the lines $x_2 = x_1$ and $x_2 = -x_1$.

**c)** This is not a Gaussian random vector, and shows that in general, marginally Gaussian random variables do not imply jointly Gaussian random variables.

**2.** For binary classification the MAP classification rule can be expressed using the log-likelihood ratio: if

$$\log\left(\frac{p(\boldsymbol{x}|y=0)}{p(\boldsymbol{x}|y=1)}\right) > 0 \tag{1}$$

then $\widehat{y} = 0$, else $\widehat{y} = 1$.

In *Gaussian discriminant analysis*, we assume that the class conditional distributions are Gaussian:

$$p(\boldsymbol{x}|y=0) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\right)$$

and

$$p(\boldsymbol{x}|y=1) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\right).$$

**a)** In linear discriminant analysis, $\mathbf{\Sigma}_0 = \mathbf{\Sigma}_1$, and (1) is equivalent to the following linear classification rule: if

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

then $\widehat{y} = 0$, else $\widehat{y} = 1$. Find $\boldsymbol{w}$ and $c$ in terms of $\mathbf{\Sigma}_0$, $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$.

**b)** In quadratic discriminant analysis, $\mathbf{\Sigma}_0 \neq \mathbf{\Sigma}_1$, and (1) is equivalent to the following quadratic rule: if

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

then $\widehat{y} = 0$, else $\widehat{y} = 1$. Find an expression for $\boldsymbol{B}$, $\boldsymbol{w}$ and $c$ in terms of $\mathbf{\Sigma}_0$, $\mathbf{\Sigma}_1$, $\mu_0$ and $\mu_1$.

**SOLUTION:**

**a)** It is sufficient to look at the general case first, and then plug in $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_0$ for the linear case. Substituting the pdfs into the log-likelihood ratio, we have

$$\log(|\mathbf{\Sigma}_1|) - \log(|\mathbf{\Sigma}_0|) - (\boldsymbol{x} - \boldsymbol{\mu}_0)^T\mathbf{\Sigma}_0^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_0) + (\boldsymbol{x} - \boldsymbol{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1) > 0$$

and after expanding and collecting terms,

$$\boldsymbol{x}^T(\mathbf{\Sigma}_1^{-1} - \mathbf{\Sigma}_0^{-1})\boldsymbol{x} + 2\boldsymbol{x}^T(\mathbf{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1) > \log(|\mathbf{\Sigma}_0|) - \log(|\mathbf{\Sigma}_1|) + \boldsymbol{\mu}_0^T\mathbf{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1^T\mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1$$

and we have

$$\begin{aligned} c &= \log(|\mathbf{\Sigma}_0|) - \log(|\mathbf{\Sigma}_1|) + \boldsymbol{\mu}_0^T\mathbf{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1^T\mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1 \\ \boldsymbol{w} &= 2(\mathbf{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1) \\ \boldsymbol{B} &= \mathbf{\Sigma}_1^{-1} - \mathbf{\Sigma}_0^{-1}. \end{aligned}$$