1. Suppose the joint probability density function of two continuous random variables X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} 6e^{-(2x+3y)} & \text{if } x > 0, y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Find the marginal probability density function f_X of random variable X.

   Ans: The marginal probability density function f_X can be obtained by integrating the joint probability density function with respect to y.

$$f_X(x) = \int_{\mathbf{R}} f_{X,Y}(x,y)\, dy = \begin{cases} \int_0^\infty 6e^{-(2x+3y)}\, dy = 2e^{-2x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

   (b) Compute the expected value $\mathbb{E}[X]$.

   Ans: Using the definition of the expectation,

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathbf{R}} x f_X(x)\, dx = \int_0^\infty x\ 2e^{-2x}\, dx \\ &= -xe^{-2x}\Big|_0^\infty + \int_0^\infty e^{-2x}\, dx = \frac{1}{2}\ , \end{aligned}$$

   where we used the integration by parts.

   (c) Determine whether or not X and Y are independent.

   Ans: Following the same steps in part (a), the marginal probability density function of Y is given by

$$f_Y(y) = \int_{\mathbf{R}} f_{X,Y}(x,y)\, dx = \begin{cases} \int_0^\infty 6e^{-(2x+3y)}\, dx = 3e^{-2y} & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

   It is clear that $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for all $x, y \in \mathbb{R}$. Hence, X and Y are independent.

2. A tumor can be either benign or malignant. The growth rate of a tumor, which we denote by X, can be modeled as a random variable, whose distribution depends on the type of tumor. The growth rate of a benign tumor has a probability density function

$$p_B(x) = \begin{cases} 5e^{-5x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

   On the other hand, the growth rate of a malignant tumor has a probability density function

$$p_M(x) = \begin{cases} 2e^{-2x} & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

A doctor measures the size of a tumor on two different dates and computes the growth rate of the tumor in a patient. Based on the measured growth rate, the doctor has to tell a patient whether the tumor is benign or malignant. When the doctor makes a correct diagnosis, there is no cost. If the doctor tells a patient that he/she has a malignant tumor when it is benign, the doctor incurs a cost of 3. On the other hand, when the doctor tells a patient that he/she has a benign tumor when it is in fact malignant, the doctor suffers a cost of 20. Suppose that only 10 percent of all tumors are malignant. Design a Bayes classifier that minimizes the overall risk.

Ans: Let X be the measured growth rate. First,

$$p(x) = 0.9p_B(x) + 0.1p_M(x) = 4.5e^{-5x} + 0.2e^{-2x} .$$

For $x \in (0, \infty)$, using the expression for the posterior probability, we can compute the following posterior probabilities.

$$\begin{aligned}
\mathbb{P}\left(\text{benign}|x\right) &= \frac{p_B(x)\mathbb{P}\left(\text{benign}\right)}{p(x)} = \frac{4.5e^{-5x}}{4.5e^{-5x} + 0.2e^{-2x}} \\
\mathbb{P}\left(\text{malignant}|x\right) &= \frac{p_M(x)\mathbb{P}\left(\text{malignant}\right)}{p(x)} = \frac{0.2e^{-2x}}{4.5e^{-5x} + 0.2e^{-2x}}
\end{aligned}$$

The conditional risks are given by the following:

$$\begin{aligned}
R(\text{malignant}|x) &= \lambda(\text{malignant}|\text{benign})\mathbb{P}\left(\text{benign}|x\right) = 3\frac{4.5e^{-5x}}{4.5e^{-5x} + 0.2e^{-2x}} \\
R(\text{benign}|x) &= \lambda(\text{benign}|\text{malignant})\mathbb{P}\left(\text{malignant}|x\right) = 20\frac{0.2e^{-2x}}{4.5e^{-5x} + 0.2e^{-2x}}
\end{aligned}$$

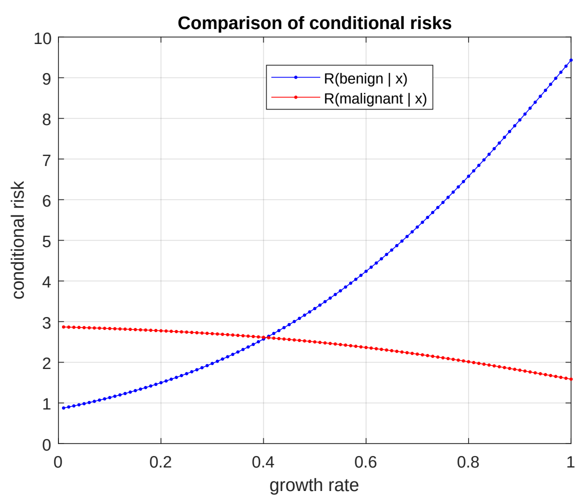


Figure 1: Comparison of conditional risks.

Figure 1 shows the two conditional risks. Based on the plot, the threshold on x is approximately 0.405. In other words, the Bayes classifier chooses ‘benign’ if the growth rate is less than 0.405 and selects ‘malignant’ otherwise.

3. The distribution of observations $\mathbf{X} = (X_1, X_2)$, which are sensor measurements we take each day to detect a possible gas leak, depends on whether or not there is a gas leak at the time. When there is a gas leak, $\mathbf{X}$ is jointly Gaussian with mean $\boldsymbol{\mu}_{GL} = [10\ 10]^T$. When there is no gas leak, $\mathbf{X}$ is jointly Gaussian with mean $\boldsymbol{\mu}_{NL} = \mathbf{0}$. The probability that there is a gas leak on any given day is 0.1.

   (a) Assume that the covariance matrix $\Sigma = \sigma^2 \mathbf{I}_{2\times 2}$ in both cases, where $\sigma = 10$. Design the Bayes classifier that selects either {gas leak} or {no gas leak}) and minimizes the probability of error, on the basis of the sensor measurements $\mathbf{X}$. Draw the two decision regions $\mathcal{R}_{GL}$ and $\mathcal{R}_{NL}$.

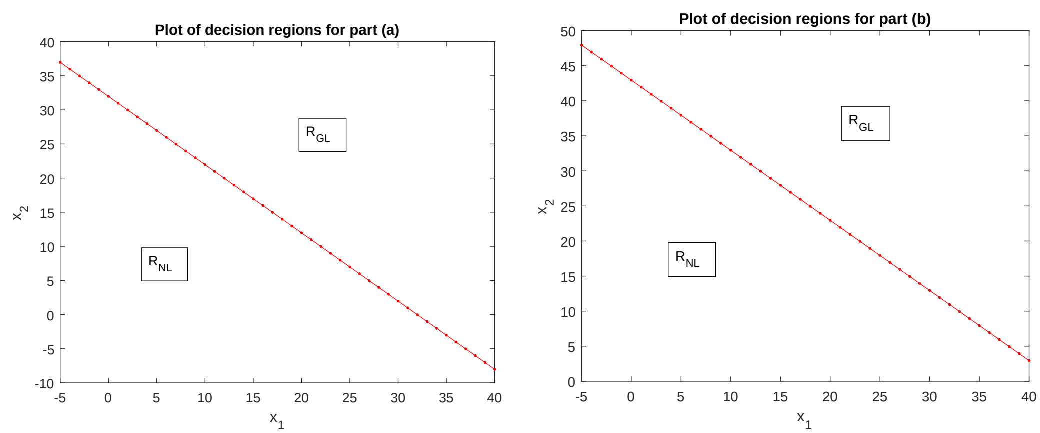
   Ans: From the lecture slide, the linear discriminant functions are given by

$$g_{GL}(\mathbf{x}) = \frac{1}{10^2}[10\ 10]^T\mathbf{x} - \frac{10^2+10^2}{2\cdot 10^2} + \log(0.1) = \frac{x_1+x_2}{10} - 3.303$$
$$g_{NL}(\mathbf{x}) = \frac{1}{10^2}[0\ 0]^T\mathbf{x} - \frac{0}{2\cdot 10^2} + \log(0.9) = -0.105$$

   The boundary of the two decision regions satisfies $g_{GL}(\mathbf{x}) = g_{NL}(\mathbf{x})$:

$$\frac{x_1+x_2}{10} - 3.303 = -0.105,$$

   or equivalently, $x_1 + x_2 = 31.98$. This is shown in Figure 2(a).



(a) (b)

Figure 2: Comparison of conditional risks.

   (b) Suppose that the covariance matrix Σ is given by

$$\Sigma = \begin{bmatrix} 100 & 50 \\ 50 & 100 \end{bmatrix}.$$

   Determine the decision regions $\mathcal{R}_{GL}$ and $\mathcal{R}_{NL}$ for the Bayes classifier.

   Ans: Following the same steps with a new covariance matrix,

$$g_{GL}(\mathbf{x}) = [0.0667\ 0.0667]^T\mathbf{x} - 0.667 - 2.303 = 0.0667(x_1+x_2) - 2.97$$
$$g_{NL}(\mathbf{x}) = -0.105$$

The boundary of the two decision regions satisfies $g_{GL}(\mathbf{x}) = g_{NL}(\mathbf{x})$:

$$0.0667(x_1 + x_2) - 2.97 = -0.105,$$

or equivalently, $x_1 + x_2 = 42.95$. This is shown in Figure 2(b).