# Bayesian Classification Example I

**This live-script file simulates the Baysian classification process within the context of a binary classification problem between classes $\omega_1$ and $\omega_2$. We asssume that the underlying feature space is one-dimensional and that the class-conditional probability distributions are given by the normal distribution parameterized by $\mu$ and $\sigma^2$.**

**Let $P_1 = P(\omega_1)$ be the a-priori probability for the first class and $P_2 = P(\omega_2)$ be the a-priori probability for the second class. Moreover, the class-conditional probability distributions will be given as follows:**

$$p(x|\omega_1;\mu_1,\sigma_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$ **[1] and**

$$p(x|\omega_2;\mu_2,\sigma_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$ **[2]**

**Bayesian classification will be conducted by identifying the corresponding decision regions in the one-dimensional feature space where the associated a-posteriori probabilities are maximimized for each class. Thus, we may write:**

$$R_1 = \{x \in \mathbb{R} : P(\omega_1|x) > P(\omega_2|x)\}$$ **[3] and**

$$R_2 = \{x \in \mathbb{R} : P(\omega_2|x) > P(\omega_1|x)\}$$ **[4]**

```
% Clear workspace and command window.
clc
clear
% Initialize the prior probabilities for each class.
P1 = 0.5;
P2 = 1-P1;
% Initialize the internal parameters for the first Gaussian distribution.
mu_1    =    1;
```

```
sigma_1 =    1;
% Initialize the internal parameters for the second Gaussian distribution.
mu_2     =  -1;
sigma_2  =   1;

% Letting (mu_j,sigma_j) for j in {1,2} to denote the internal parameters
% of the Gaussian distributions for the two classes, we may define an extra
% variable k that parameterizes the interval in which each class-conditional
% probability density function is defined. In this context, we may define the
% interval for each class-conditional pdf as:
% X_j = [mu_j - k * sigma_j,mu_j + k * sigma_j] for j in {1,2}.
% In this context, the one-dimensional feaure space X may be defined as:
% X  =  [mu_min - k * sigma_max,mu_max + k * sigma_max]
k = 3;
mu_min = min(mu_1,mu_2);
mu_max = max(mu_1,mu_2);
sigma_max = max(sigma_1,sigma_2);
% Set the dx parameter controlling the x-step.
dx = 0.001;
% Set the dy parameter controlling the y-step.
dy = 0.001;
% Define the one-dimentional feature space X.
X = mu_min-k*sigma_max:dx:mu_max+k*sigma_max;
% Compute the class-conditional probability distribution functions for the
% classes p_x_w1 and p_x_w2 within the previously defined interval.
p_x_w1 = normpdf(X,mu_1,sigma_1);
p_x_w2 = normpdf(X,mu_2,sigma_2);
% Compute the posterior probabilities for each class.
P_w1_x = p_x_w1 * P1;
P_w2_x = p_x_w2 * P2;
% Determine the minimum and maximum values of the X-axis.
Xmin = min(X);
Xmax = max(X);
% Determine the minimum and maximum values of the Y-axis.
Ymin = min([min(P_w1_x),min(P_w2_x)]);
Ymax = max([max(P_w1_x),max(P_w2_x)]);
% Define the Y-axis.
Y = Ymin:dy:Ymax;
% Plot the posterior probabilities for each class.
figure('Name','Bayesian Classfication: Posterior Probablities');
title('Bayesian Classfication: Posterior Class Probablities');
hold on
plot(X,P_w1_x,'-r','Linewidth',2.5);
plot(X,P_w2_x,'-b','LineWidth',2.5);
xlabel('x');
ylabel('P(w|x)');
% Shade the area below each probability density function.
H1=area(X,P_w1_x,'FaceColor','r');
H1.FaceAlpha = 0.2;
H2=area(X,P_w2_x,'FaceColor','b');
H2.FaceAlpha = 0.2;
```

**In this framework, the Bayesian classification rule may be formulated by the following discrimination function:**

$$g(x) = \begin{cases} \omega_1, & x \in R_1 \\ \omega_2, & x \in R_2 \end{cases} \quad \textbf{[5]}$$

**Therefore, the decision boundary** $R_0 = \{x \in \mathbb{R} : h(x) = P(\omega_1|x) - P(\omega_2|x) = 0\}$ **[6] between the classification regions** $R_1$ **and** $R_2$ **may be computed by solving the following equation:**

$$P(\omega_1|x) = P(\omega_2|x) \Rightarrow$$

$$\frac{p(x|\omega_1)P(\omega_1)}{p(x)} = \frac{p(x|\omega_2)P(\omega_2)}{p(x)} \Rightarrow$$

$$\frac{p(x|\omega 1)}{p(x|\omega_2)} = \frac{P(\omega_2)}{P(\omega_1)} \Rightarrow$$

$$e^{\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}} = \frac{P(\omega_2)}{P(\omega_1)} \frac{\sigma_1}{\sigma_2} \quad \textbf{[7]}$$

**Letting** $z = \dfrac{(x-\mu_2)^2}{2\sigma_2^2} - \dfrac{(x-\mu_1)^2}{2\sigma_1^2}$ **[8], we may write that**

$$z = ln\frac{P(\omega_2)}{P(\omega_1)} + ln\frac{\sigma_1}{\sigma_2} \quad \textbf{[9]}$$

**Moreover, by letting** $\lambda_p = ln\frac{P(\omega_2)}{P(\omega_1)}$ **[10] and** $\lambda_\sigma = ln\frac{\sigma_1}{\sigma_2}$ **[11], Equation (6) may be re-written as:**

$$\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2} = \lambda_p + \lambda_\sigma \textbf{ [12]}$$

**Equation (12) suggests that the decision boundary between classes** $\omega_1$ **and** $\omega_2$ **will be given by the solutions of the following equation:**

$$h(x) = (\sigma_1^2 - \sigma_2^2)x^2 + 2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2)x + \sigma_1^2\mu_2^2 - \sigma_2^2\mu_1^2 - 2\sigma_1^2\sigma_2^2(\lambda_p + \lambda_\sigma) = 0$$

**[13]**

**By letting** $\alpha = (\sigma_1^2 - \sigma_2^2)$ **[14],** $\beta = 2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2)$ **[15] and** $\gamma = \sigma_1^2\mu_2^2 - \sigma_2^2\mu_1^2 - 2\sigma_1^2\sigma_2^2(\lambda_p + \lambda_\sigma)$ **[16], we may write that** $h(x) = \alpha x^2 + \beta x + \gamma$ **[17]. Thus, the discriminant for the second degree polynomial** $h(x)$ **can be expresse as:** $\Delta_h = \beta^2 - 4\alpha\gamma$**.**

**In this context, we may identify the following cases for the decision boundary** $R_0 = \{x \in \mathbb{R} : h(x) = 0\}$ **[18].**

**Case I:** $\alpha = 0, \beta \neq 0 \Rightarrow R_0 = \{x_0\} = \{-\frac{\gamma}{\beta}\}$ **(**$h(x) = \beta x + \gamma$**)**

**Case II:** $\alpha \neq 0, \Delta_h > 0 \Rightarrow R_0 = \{x_0^+, x_0^-\} = \{\frac{-\beta + \sqrt{\Delta_h}}{2\alpha}, \frac{-\beta - \sqrt{\Delta_h}}{2\alpha}\}$

**(**$h(x) = \alpha(x - x_0^-)(x - x_0^+)$**)**

**In fact, we set that** $x_0^+ = \max\{\frac{-\beta + \sqrt{\Delta_h}}{2\alpha}, \frac{-\beta - \sqrt{\Delta_h}}{2\alpha}\}$ **and**

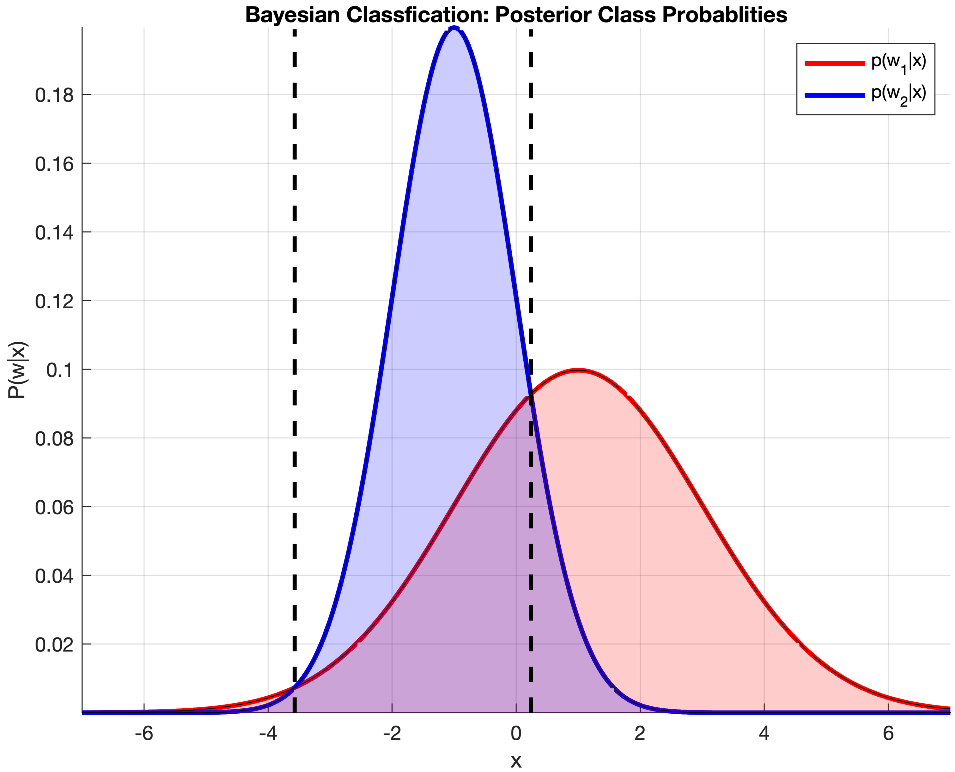
$$x_0^- = \min\{\frac{-\beta + \sqrt{\Delta_h}}{2\alpha}, \frac{-\beta - \sqrt{\Delta_h}}{2\alpha}\}$$

**Case III:** $\alpha \neq 0, \Delta_h = 0 \Rightarrow R_0 = \{\hat{x}_0\} = \{-\frac{\beta}{2\alpha}\}$ **(**$h(x) = \alpha(x - \hat{x}_0)^2$**)**

**Case IV:** $\alpha \neq 0, \Delta_h < 0 \Rightarrow R_0 = \{\varnothing\}$

**Cases III and IV will not be considered since they involve occasions where we have either** $h(x) \geq 0, \forall x \in \mathbb{R}$ **or** $h(x) \leq 0, \forall x \in \mathbb{R}$ **respectively, which, in turn, result in the following degenerate cases** $R_2 = \{\varnothing\}$ **or** $R_1 = \{\varnothing\}$**.**

```
% Compute the constant values Lp and Ls.
Lp = log(P2/P1);
Ls = log(sigma_1/sigma_2);
% Compute the coefficients of the polynomial h(x) = A*x^2 + B*x + C whose roots
% determine the boundary region Ro between the two classes.
A = sigma_1^2 - sigma_2^2;
B = 2*(mu_1*sigma_2^2-mu_2*sigma_1^2);
C = (sigma_1*mu_2)^2  - (sigma_2*mu_1)^2 - 2*((sigma_1*sigma_2)^2)*(Lp+Ls);
% Compute the discriminant D of the polynomial h(x).
D = B^2 - 4*A*C;
% Compute the solution points for the quadratic equation: h(x) = 0 and plot
% the corresponding boundary.
if(A==0)
    Xo = - C/B;
    plot(Xo*ones(1,length(Y)),Y,'--k','LineWidth',2.0);
else
    if(D>0)
        Xo_plus = max(((-B + sqrt(D)) / (2*A)),((-B - sqrt(D)) / (2*A)));
        Xo_minus = min(((-B + sqrt(D)) / (2*A)),((-B - sqrt(D)) / (2*A)));
        plot(Xo_plus*ones(1,length(Y)),Y,'--k','LineWidth',2.0);
        plot(Xo_minus*ones(1,length(Y)),Y,'--k','LineWidth',2.0);
    end
end
hold off
legend({'p(w_1|x)','p(w_2|x)'});
grid on
axis([Xmin Xmax Ymin Ymax]);
```




**According to the previous analysis, decision regions $R_1$ and $R_2$ may be reformulated as:** $R_1 = \{x \in \mathbb{R} : h(x) > 0\}$ **[19] and** $R_2 = \{x \in \mathbb{R} : h(x) < 0\}$ **[20]**

**Thus, we may consider the following cases for the decision regions $R_1$ and $R_2$:**

**Case I:** $\alpha = 0, \beta \neq 0 \Rightarrow R_1 = (x_0, +\infty)\ \text{ and }\ R_2 = (-\infty, x_0)$

**Case IIa:** $\alpha > 0, \Delta_h > 0 \Rightarrow R_1 = (-\infty, x_0^-) \cup (x_0^+, +\infty)\ \text{ and }\ R_2 = (x_0^-, x_0^+)$

**Case IIb:** $\alpha < 0, \Delta_h > 0 \Rightarrow R_1 = (x_0^-, x_0^+)\ \text{ and }\ R_2 = (-\infty, x_0^-) \cup (x_0^+, +\infty)$

**In this setting, the overall probability of error is given by the following formula:**

$$P(error) = P(\omega_1) \int_{R_2} p(x|\omega_1)dx + P(\omega_2) \int_{R_1} p(x|\omega_2)dx \ \ [22]$$

**Case I: Equation (22) may be rewritten in the following form:**

$$P(error) = P(\omega_1)\int_{-\infty}^{x_0} p(x|\omega_1)dx + P(\omega_2)\int_{x_0}^{+\infty} p(x|\omega_2)dx \quad [23]$$

**Case IIa: Equation (22) may be rewritten in the following form:**

$$P(error) = P(\omega_1)\int_{x_0^-}^{x_0^+} p(x|\omega_1)dx + P(\omega_2)\left[\int_{-\infty}^{x_0^-} p(x|\omega_2)dx + \int_{x_0^+}^{+\infty} p(x|\omega_2)dx\right] \quad [24]$$

**Case IIb: Equation (22) may be rewritten in the following form:**

$$P(error) = P(\omega_1)\left[\int_{-\infty}^{x_0^-} p(x|\omega_1)dx + \int_{x_0^+}^{+\infty} p(x|\omega_1)dx\right] + P(\omega_2)\int_{x_0^-}^{x_0^+} p(x|\omega_2)dx \quad [25]$$

**Evaluating the integrals appearing within Equations (23), (24) and (25) requires the existence of a computational procedure for estimating the cumulative distribution function that corresponds to a given probability distribution function. For the case of a normal probability distribution function, the associated cumulative distribution function is given by:**

$$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = I(x;\mu,\sigma) = normcdf(x, mu, sigma) \quad [26]$$

**Therefore, we may define the following auxiliary integrals as:**

$$\int_{x}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = 1 - I(x;\mu,\sigma) \quad [27]$$

**and**

$$\int_{x_a}^{x_b} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = I(x_b;\mu,\sigma) - I(x_a;\mu,\sigma) \quad [28]$$

**In view of Equations (26), (27) and (28) the overall error probabilities for the previously identified cases may be given as:**

**Case I:** $P(error) = P(\omega_1)I(x_0; \mu_1, \sigma_1) + P(\omega_2)[1 - I(x_0; \mu_2, \sigma_2)]$ **[29]**

**Case IIa:**

$P(error) = P(\omega_1)[I(x_0^+; \mu_1, \sigma_1) - I(x_0^-; \mu_1, \sigma_1)] + P(\omega_2)[I(x_0^-; \mu_2.\sigma_2) + 1 - I(x_0^+; \mu_2$

**[30]**

**Case IIb:** $P(error) =$

$P(\omega_1)[I(x_0^-; \mu_1.\sigma_1) + 1 - I(x_0^+; \mu_1.\sigma_1)] + P(\omega_2)[I(x_0^+; \mu_2, \sigma_2) - I(x_0^-; \mu_2, \sigma_2)]$

**[31]**

```
% Compute the overall error probabiliities for each one of the previously
% identified cases.
if(A==0)
    Pe = P1 * normcdf(Xo,mu_1,sigma_1) + P2 * (1  - normcdf(Xo,mu_2,sigma_2))
else
    if (D > 0)
        if (A > 0)
            Pe = P1 * (normcdf(Xo_plus,mu_1,sigma_1) - normcdf(Xo_minus,mu_1,sigma_1))
        else
            Pe = P1 * (normcdf(Xo_minus,mu_1,sigma_1) + 1 - normcdf(Xo_plus,mu_1,sigma_
        end
    end
end
```

```
Pe = 0.2267
```