Machine Learning I at TU Berlin

Assignment 1 - Group PTHGL

October 21, 2018

Exercise 1

(a)

Given equations:

$$
\begin{aligned}
(I) \qquad & P(\text{error}|x) = min[P(w_1|x), P(w_2|x)] \\
(II) \qquad & P(\text{error}) = \int P(\text{error}|x)\ p(x)\mathrm{d}x \\
(III) \qquad & P(\text{error}) \leq \int \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}}\ p(x)\mathrm{d}x \\
& = \int \frac{2P(w_1|x)P(w_2|x)}{P(w_1|x) + P(w_2|x)}\ p(x)\mathrm{d}x \qquad /P(w_1|x) + P(w_2|x) = 1 \\
& = \int 2P(w_1|x)P(w_2|x)\ p(x)\mathrm{d}x
\end{aligned}
$$

Assume:

$$P(w_1|x) \geq P(w_2|x) \Rightarrow min[P(w_1|x), P(w_2|x)] = P(w_2|x)$$

Thus (I) and (II) can be transposed as followed:

$$
\begin{aligned}
P(\text{error}|x) &= P(w_2|x) \\
P(\text{error}) &= \int P(\text{error}|x)\ p(x)\mathrm{d}x = \int P(w_2|x)\ p(x)\mathrm{d}x
\end{aligned}
$$

Since:

$$0 \leq P(w_1|x) \leq 1, \quad 0 \leq P(w_2|x) \leq 1, \quad P(w_1|x) = 1 - P(w_2|x)$$

We can say:

$$P(w_2|x) \geq P(w_1|x)P(w_2|x)$$

and:

$$P(w_1|x) \geq \frac{1}{2} \quad \Rightarrow \quad P(w_2|x) \leq 2P(w_1|x)P(w_2|x)$$

Thus our new $P(\text{error})$ is:

$$\begin{aligned}
P(\text{error}) \leq \int 2P(w_1|x)P(w_2|x)\ p(x)\mathrm{d}x &= \int \frac{2P(w_1|x)P(w_2|x)}{1}\ p(x)\mathrm{d}x \\
&= \int \frac{2P(w_1|x)P(w_2|x)}{P(w_1|x)+P(w_2|x)}\ p(x)\mathrm{d}x \\
&= \int \frac{2}{\frac{1}{P(w_1|x)}+\frac{1}{P(w_2|x)}}\ p(x)\mathrm{d}x
\end{aligned}$$

(b)

We use the previously calculated error probability:

$$\begin{aligned}
P(\text{error}) &\leq \int \frac{2}{\frac{1}{P(w_1|x)}+\frac{1}{P(w_2|x)}}\ p(x)\mathrm{d}x \\
&\leq \int \frac{2P(w_1|x)\cdot P(w_2|x)}{P(w_1|x)+P(w_2|x)}\ p(x)\mathrm{d}x && /P(w_j|x) = \frac{p(x|w_j)P(w_j)}{p(x)} \\
&\leq \int \frac{2p(x|w_1)P_1\cdot p(x|w_2)P_2}{p(x|w_1)P_1+p(x|w_2)P_2}\mathrm{d}x && /P_j := P(w_j)
\end{aligned}$$

Insert the given univariate probability distribution $p(x|w_1)$ and $p(x|w_2)$.

$$\begin{aligned}
P(\text{error}) &\leq \frac{2P_1P_2}{\pi}\int \frac{1}{(P_1+P_2)x^2+2\mu(P_1-P_2)x+(P_1+P_2)(1+\mu^2)}\mathrm{d}x && /P_1+P_2=1 \\
&= \frac{2P_1P_2}{\pi}\int \frac{1}{x^2+2\mu(P_1-P_2)x+(1+\mu^2)}\mathrm{d}x && /\text{use identity} \\
&= \frac{2P_1P_2}{\pi}\frac{2\pi}{\sqrt{4(1+\mu^2)-4\mu^2(P_1-P_2)^2}} \\
&= \frac{2P_1P_2}{\sqrt{1+\mu^2(1-P_1^2+2P_1P_2-P_2^2)}} \overset{*}{=} \frac{2P_1P_2}{\sqrt{1+4\mu^2P_1P_2}}
\end{aligned}$$

(*) $(P_1+P_2)^2 = 1^2 \;\Leftrightarrow\; P_1^2 = 1-2P_1P_2-P_2^2.$

(c)

If the data is low-dimensional, we can find the error by calculating the overlapping area(1D) or volume(2D) within the smaller probability density functions, and integrate over the borders. For low-dimensional data, the calculation is fairly easy. The higher the dimension, the more complex and time consuming this calculation gets and another approach would be necessary.

For high-dimensional data, one option would be to downsize the data by sampling and using the filtered data to calculate an average error. One filter could be by clustering, or random sampling, nearest mean filter and others.

Exercise 2

(a)

We define the following discriminant function:

$$\begin{aligned} g(x) \equiv g_1(x) - g_2(x) &= P(w_1|x) - P(w_2|x) \\ &= \frac{p(x|w_1)P(w_1) - p(x|w_2)P(w_2)}{p(x)} \qquad / \cdot p(x) \\ \Leftrightarrow \qquad &= p(x|w_1)P(w_1) - p(x|w_2)P(w_2) \end{aligned}$$

We can multiply multiply the discriminant functions by the same positive constant or shift them by the same additive constant without influencing the decision. Applying a monotonically function to the discriminant function does not change the decision rule either. Hence, if we apply the natural logarithmic function to our discriminant function we get

$$g(x) = \ln\left[\frac{p(x|w_1)}{p(x|w_2)}\right] + \ln\left[\frac{P(w_1)}{P(w_2)}\right] \quad .$$

Insert the univariate Laplacian probability distribution and we get

$$g(x) = \frac{|x+\mu| - |x-\mu|}{\sigma} + \ln\left[\frac{P(w_1)}{P(w_2)}\right] \qquad , \mu, \sigma > 0 \quad .$$

We get the optimal decision boundary if we set $g(x)$ equal to zero and solve for x_b

$$\begin{aligned} 0 &= \frac{|x_b+\mu| - |x_b-\mu|}{\sigma} + \ln\left[\frac{P(w_1)}{P(w_2)}\right] \\ \Rightarrow x_b &= \pm \frac{s\sqrt{s^2-4\mu^2}}{2\sqrt{s^2-4\mu^2}} \qquad \text{with} \quad s^2 - 4\mu^2 \geq 0, s = -\ln\left[\frac{P(w_1)}{P(w_2)}\right]\sigma \end{aligned}$$

(b)

With the discriminant function that we defined we can use the following decision rule:

Decide w_1 if $g(x) > 0$ else decide w_2.

In the following cases the optimal decision always predict w_1:

Case 1:

$$\mu > 0, \quad \sigma \to \infty, \quad \frac{P(w_1)}{P(w_2)} > 1$$

When $\sigma \to \infty$, the Laplacian probability distribution nears a uniform distribution. Therefore μ doesn't influence the decision. The ratio of $P(w_1)$ and $P(w_2)$ now decides the outcome.

Case 2:

$$\mu > 0, \quad \sigma > 0, \quad \frac{P(w_1)}{P(w_2)} > 1, \quad \forall x : \left|\frac{|x+\mu| - |x-\mu|}{\sigma}\right| > \ln\left[\frac{P(w_1)}{P(w_2)}\right]$$

Case 2 is always true for any given x if there is no occurrence of w_2, i.e. $P(w_2) \to 0$.

(c)

For the given univariate Gaussian probability distributions we get this discriminant function

$$g(x) = \frac{2x\mu}{\sigma^2} + \ln\left[\frac{P(w_1)}{P(w_2)}\right] \qquad , \sigma > 0 \quad .$$

We get the optimal decision boundary if we set $g(x)$ equal to zero and solve for the boundary x_b

$$x_b = -\frac{\sigma^2}{2\mu}\ln\left[\frac{P(w_1)}{P(w_2)}\right] \quad .$$

In the following cases the optimal decision always predict w_1:

Case 1:

$$\mu > 0, \quad \sigma \to \infty, \quad P(w_1) > P(w_2)$$

When $\sigma \to \infty$, the Gaussian probability distribution nears a uniform distribution. Therefore μ doesn't influence the decision and $P(w_1)$ needs to be larger than $P(w_2)$, to always decide for w_1 (equal to case 1 for Laplacian distribution).

Case 2:

$$P(w_1) > P(w_2), \quad \forall x, \mu, \sigma : \ln\left[\frac{P(w_1)}{P(w_2)}\right] > \left|\frac{2x\mu}{\sigma^2}\right|$$

Assuming $\frac{2x\mu}{\sigma^2}$ can not be infinte, the second case can only be garanteed for any x, μ and σ, when $P(w_2) \to 0$ (equal to case 2 for Laplacian distribution).