CS123a Statistical Machine Learning (Spring 2013): Homework Assignment #2

Chuan Wang

February 1, 2013

Problem 1

1.1

(1)

When $f_1(t) = sell$,

$$\begin{aligned}R(f_1) =& P(w_H) * \lambda(sell \mid w_H) + P(w_S) * \lambda(sell \mid w_S) \\ =& -0.99 * \$1 + 0.01 * \$500 = \$4.01\end{aligned}$$

(2)

When $f_2(t) = kill$,

$$\begin{aligned}R(f_1) =& P(w_H) * \lambda(kill \mid w_H) + P(w_S) * \lambda(kill \mid w_S) \\ =& -0.99 * \$0.5 + 0.01 * \$0.5 = \$0.5\end{aligned}$$

(3)

When $f_3(t_{Normal}) = sell, f_3(t_{High}) = kill$,

$$\begin{aligned}P(t_N) =& \sum_{w_H, w_S} P(t_N \mid w) = 0.9802 \\ P(t_H) =& 1 - P(t_N) = 0.0198 \\ P(w_H \mid t_N) =& \frac{P(t_N \mid w_H)P(w_H)}{P(t_N)} = \frac{0.99 * 0.99}{0.9802} = 0.9999 \\ P(w_S \mid t_H) =& 1 - P(w_H \mid t_H) = 0.0001 \\ P(w_H \mid t_H) =& \frac{P(t_H \mid w_H)P(w_H)}{P(t_H)} = \frac{0.01 * 0.99}{0.0198} = 0.5 \\ P(w_S \mid t_H) =& 1 - P(w_H \mid t_H) = 0.5\end{aligned}$$

So the total risk of the decision function is:

$$
\begin{aligned}
E_{P(t)}[R(f_3(t) \mid t)] = & \sum_t P(t) \cdot R(f_3 \mid t) \\
= & P(t_N)(P(w_H \mid t_N) * \lambda(sell \mid w_H) \\
& + P(w_S \mid t_N) * \lambda(sell \mid w_S)) \\
& + P(t_H)(P(w_H \mid t_H) * \lambda(kill \mid w_H) \\
& + P(w_S \mid t_H) * \lambda(kill \mid w_S)) \\
= & -0.9212
\end{aligned}
$$

1.2

To make the industry to be profitable, the total risk should be negative.So,

$$
\begin{aligned}
E_{P(t)}[R(f_3(t) \mid t)] = & P(t_N)(P(w_H \mid t_N) * \lambda(sell \mid w_H) \\
& + P(w_S \mid t_N) * \lambda(sell \mid w_S)) \\
& + P(t_H)(P(w_H \mid t_H) * \lambda(kill \mid w_H) \\
& + P(w_S \mid t_H) * \lambda(kill \mid w_S)) < 0 \\
\lambda(sell \mid w_S) < 9898 &
\end{aligned}
$$

So the upper limit of $\lambda(sell \mid w_S)$ is \$9898

Problem 2

(2.1)

There are two kinds of errors: (1) Classify $\vec{x}$ into class 1 if $x \leq \tau_0$,(2) Otherwise classify $\vec{x}$ into class 2.

$$
\begin{aligned}
P(error) = & P(x \in w_1, x \leq \tau_0) + P(x \in w_2, x > \tau_0) \\
= & P(w_1)P(x \leq \tau_0 \mid w_1) + P(w_2)P(x > \tau_0 \mid w_2) \\
= & P(w_1) \int_{-\infty}^{\tau_0} p(x \mid w_1)\mathrm{d}x + P(w_2) \int_{\tau_0}^{\infty} p(x \mid w_2)\mathrm{d}x
\end{aligned}
$$

2.2

To minimize the $P(error)$, the necessary condition is that first derivative should be zero.
By differentiating the $P(error)$ with τ_0, we get $P(w_1)p(\tau_0) \mid w_1) - P(w_2)p(\tau_0 \mid w_2) = 0$. So it is proved.

2.3

As shown in Figure 1 the decision boundary $x^* = \tau_0$ must satisfy the condition in (2.2), the x^* actually gives the minimum of the $p(error)$.However if we consider

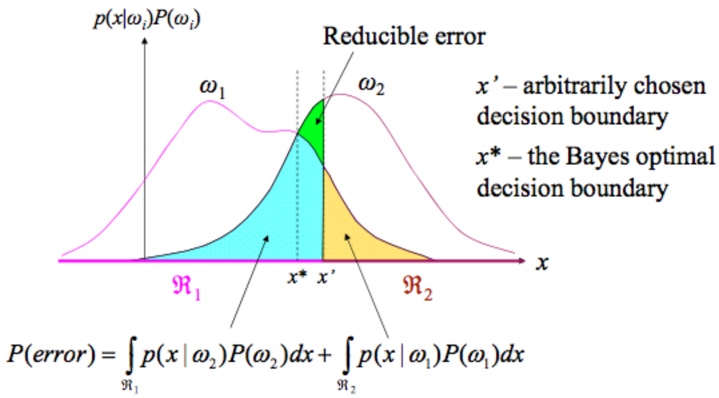

Figure 1: Error Probability

the opposite way, the left side area of the decision boundary under the pdf curve corresponding to $P(x \in w_1, x \leq \tau_0)$ and right side area corresponding to $P(x \in w_2, x > \tau_0)$. And we can see that x^* gives the maximum of $p(error)$. And arbitrarily chosen $x^{'}$ will reduce the area of $p(error)$ by the area of reducible error.

Problem 3

3.1

We have the conditional risk for action a_i that is classifying x into i^{th} class:

$$\begin{aligned} R(a_i \mid x) =& \sum_{j=1}^{c} \lambda(a_i \mid w_j) P(w_j \mid x) \\ =& \sum_{j=1, j \neq i}^{c} \lambda_s P(w_j \mid x) \\ =& \lambda_s (1 - P(w_i \mid x)) \end{aligned}$$

Cause $P(w_i \mid x) \geq P(w_j \mid x)$ for all j, So

$$R(a_i \mid x) = \lambda_s(1 - P(w_i \mid x)) \leq \lambda_s(1 - P(w_j \mid x)) = R(a_j \mid x)$$

For classes that are unrecognizable, say k, the conditional risk is:

$$R(a_k \mid x) = \sum_{j=1}^{c} \lambda_0 P(w_j \mid x) = \lambda_0$$

Also, we have:

$$\begin{aligned} R(a_i \mid x) =& \lambda_s(1 - P(w_i \mid x)) \\ \leq& \lambda_s(1 - (1 - \frac{\lambda_0}{\lambda_s})) = \lambda_s = R(a_k \mid x) \end{aligned}$$

So that action a_i minimize the risk.

(3.2)

When we decide to classify x into i^{th} class, the risk should be minimum,so we have:

$$\begin{aligned} R(a_i \mid x) \leq& R(a_j \mid x) \\ \implies \lambda_s(1 - P(w_i \mid x)) \leq& \lambda_s(1 - P(w_j \mid x)) \\ \implies \lambda_s(p(x) - p(x \mid w_i)P(w_i)) \leq& \lambda_s(p(x) - p(x \mid w_j)P(w_j)) \quad \text{Using Bayes Rule} \\ \implies p(x \mid w_i)P(w_i) \geq& p(x \mid w_j)P(w_j) \\ \implies g_i(x) \geq& g_j(x) \end{aligned}$$

Also, for unrecognizable class k,

$$\begin{aligned} R(a_i \mid x) \leq& R(a_k \mid x) \\ \implies \lambda_s(1 - P(w_i \mid x)) \leq& \lambda_0 \\ \implies -P(w_i \mid x) \leq& \frac{\lambda_0 - \lambda_s}{\lambda_s} \\ \implies p(x)P(w_i \mid x) \geq& \frac{\lambda_s - \lambda_0}{\lambda_s} p(x) \\ \implies p(x \mid w_i)P(w_i) \geq& \frac{\lambda_s - \lambda_0}{\lambda_s} \sum_{j=1}^{c} P(x, w_j) \quad \text{Using Bayes Rule} \\ \implies p(x \mid w_i)P(w_i) \geq& \frac{\lambda_s - \lambda_0}{\lambda_s} \sum_{j=1}^{c} P(x \mid w_j)P(w_j) \\ \implies g_i(x) \geq& g_k(x) \end{aligned}$$

It follows maximal discrimination.So the discriminant functions are optimal.

(3.3)

Since λ_0, λ_s should not be negative, when $\lambda_0 = 0$, $R(a_i \mid x) > 0, R(a_k \mid x) = 0$, so unrecognizable class k(reject) will always be chosen and $a_i (i = 1, 2, \ldots, c)$ will never be chosen.

(3.4)

Similiar with (3.3), because $\lambda_0 > \lambda_s$, $R(a_i \mid x) = \lambda_s(1 - P(w_i \mid x)) < \lambda_0 = R(a_k \mid x)$, so reject will never be chosen.

(3.5)

From the beginning, all the decision will be choosing to reject. As $\frac{\lambda_0}{\lambda_s}$ increase and becomes greater than $1 - P(w_i \mid x)$, all the decision will be choosing one of the categories within 1 to c.

1 Problem 4

4.1

$$
\begin{aligned}
H(x) =& -\int p(x)\log p(x)\mathrm{d}x \\
=& -\int \frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2](-\log((2\pi)^{\frac{1}{2}}\sigma) - \frac{1}{2}(\frac{x-\mu}{\sigma})^2)\mathrm{d}x \\
=& \int \frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]\log((2\pi)^{\frac{1}{2}}\sigma)\mathrm{d}x + \int \frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]\frac{1}{2}(\frac{x-\mu}{\sigma})^2\mathrm{d}x \\
=& \log((2\pi)^{\frac{1}{2}}\sigma)\int p(x)\mathrm{d}x + \frac{1}{2\sigma^2}\int (x-\mu)^2 p(x)\mathrm{d}x \\
=& \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}E[(x-\mu)^2] \\
=& \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2}
\end{aligned}
$$

4.2

$$
\begin{aligned}
H(x) =& -\int p(\vec{x})[-\frac{1}{2}(\vec{x}-\mu)^\top\Lambda^{-1}(\vec{x}-\mu) - \ln(\sqrt{2\pi})^D|\Lambda|^{\frac{1}{2}}]\mathrm{d}x \\
=& \frac{1}{2}E[\sum_{ij}(x_i-\mu_i)(\Lambda^{-1})_{ij}(x_j-\mu_j)] + \frac{1}{2}\ln(2\pi)^D|\Lambda| \\
=& \frac{1}{2}E[\sum_{ij}(x_i-\mu_i)(x_j-\mu_j)(\Lambda^{-1})_{ij}] + \frac{1}{2}\ln(2\pi)^D|\Lambda| \\
=& \frac{1}{2}\sum_{ij}E[(x_i-\mu_i)(x_j-\mu_j)](\Lambda^{-1})_{ij} + \frac{1}{2}\ln(2\pi)^D|\Lambda| \\
=& \frac{1}{2}\sum_{j}\Lambda_{ij}(\Lambda^{-1})_{ij} + \frac{1}{2}\ln(2\pi)^D|\Lambda| \\
=& \frac{1}{2}\sum_{j}\mathbf{I}_{jj} + \frac{1}{2}\ln(2\pi)^D|\Lambda| \\
=& \frac{D}{2} + \frac{1}{2}\ln(2\pi)^D|\Lambda|
\end{aligned}
$$