Homework Assignment #1  
CMPUT 466/551 Machine Learning  

**Due on January 25**  
Assigned TA: Chonghai Wang  
Part weights are mentioned in %  

**P1.** Expected prediction error (EPE) is defined as:  
\[ EPE(f(x)) = \int L(y, f(x)) \Pr(dx, dy), \]  
where \( L(.,.) \) is the loss function as mentioned in [HTF] section 2.4.  
Show that  

5% (a) for \( L(y, f(x)) = (y - f(x))^2 \), minimization of EPE leads to conditional mean:  
\[
\hat{f}(x) = \arg\min_{f(x)} EPE(f(x)) = E(Y \mid X = x), \quad \text{and}
\]

15% (b) for \( L(y, f(x)) = |y - f(x)| \), minimization of EPE leads to conditional median:  
\[
\hat{f}(x) = \arg\min_{f(x)} EPE(f(x)) = \text{Median}(Y \mid X = x).
\]

[Hint: you might need Leibnitz rule for (b):  
\[
\frac{\partial}{\partial z} \int f(x, z) dx = \int \frac{\partial f}{\partial z} dx + f(b(z), z) \frac{db(z)}{dz} - f(a(z), z) \frac{da(z)}{dz}
\]

[Note: \( c \) is a median for a probability density function (pdf) \( p(x) \) when \( \int_{-\infty}^{c} p(x) dx = \int_{c}^{\infty} p(x) dx \), i.e., when \( c \) divides the area under the pdf into two equal halves.]  

**P2.**  
15% (a) For any two functions \( f_1(x) \) and \( f_2(x) \) show that  
\[
\min_x f_1(x) + \min_x f_2(x) \leq \min_x (f_1(x) + f_2(x)).
\]

15% (b) Let Risk be as defined in the context of minimax criterion (see lecture note):  
\[
\text{Risk}(P_1, R_1) = C_{22} + (C_{12} - C_{22}) \int_{R_i} p(x \mid H_2) dx 
\]
\[
+ P_1[(C_{11} - C_{22}) + (C_{21} - C_{11})] \int_{R_i} p(x \mid H_1) dx - (C_{12} - C_{22}) \int_{R_i} p(x \mid H_2) dx.
\]

Let \( g(P_1) = \min_{R_1} [\text{Risk}(P_1, R_1)] \). Prove using part (a) that \( g(P_1) \) is concave down.  

[Note: a function \( f(x) \) is concave down if for any \( x_1 \) and \( x_2 \),  
\[
\frac{f(x_1) + f(x_2)}{2} \leq f\left(\frac{x_1 + x_2}{2}\right).
\]

10% (c) Let \( g(P_1) = \min_{R_1} [\text{Risk}(P_1, R_1)] \), and \( R_1^a = \arg\min_{R_1} \text{Risk}(a, R_1) \), where \( a \) is a fixed number (prior probability). Explain why \( \text{Risk}(P_1, R_1^a) \) will be tangent to \( g(P_1) \), assuming the latter as a smooth curve.
(5%) **P3.** Consider a sensor that outputs signal \( x \) as follows. In the absence of an object (call this hypothesis \( H_1 \)), the output \( x \) follows a Gaussian distribution with mean \( \mu_1 \) and standard deviation \( \sigma_1 \). In presence of an object (call this hypothesis \( H_2 \)) the output \( x \) follows another Gaussian distribution with mean \( \mu_2 \) and standard deviation \( \sigma_2 \). Assume \( \mu_2 > \mu_1 \). Assume prior probabilities as \( P(H_1) = \frac{2}{3} \) and \( P(H_2) = \frac{1}{3} \). Compute Bayes minimum rate classifier.

[Note: The classifier is here a threshold \( T \), such as no object if \( x \leq T \), else object]

**P4.** Consider the sensor described in **P3**, except that now we do not know the prior probabilities. Assuming a 0-1 loss criterion write Risk as defined in **P2(b)** (or in the lecture note) in the linear form: \( \text{Risk}(T, P) = a(T) + P_1 b(T) \), where \( a(T) \) and \( b(T) \) are two functions of \( T \). \( T \) is the threshold of a threshold based classifier (such as no object if output signal \( x \leq T \), else object).

(15%) (a) Find out the minimax classifier from the equation \( b(T) = 0 \).

(10%) (b) Find out an expression for minimax risk in terms of error function \( \text{erf} \).

[Note: \( \text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \)]

(10%) (c) Can you find out an expression for the prior probability \( P_1 \) for which the minimax solution is obtained?
1 Problem 1

(a) \( EPE(f(x)) = \int \int (y - f(x))^2 P(x, y)dydx = \int P(x) \left[ \int (y - f(x))^2 P(y|x)dy \right] dx \)

To minimize \( EPE(f(x)) \) Let \( F(c) = \int (y-c)^2 P(y|x)dy \) when \( c=f(x) \) for a fixed \( x \)

\[ F'(c) = \frac{d}{dc} \int (y-c)^2 P(y|x)dy = -2 \int (y-c) P(y|x)dy = 0 \Rightarrow c = \int y P(y|x)dy = E(Y|X=x) \]

(b) \( EPE(f(x)) = \int P(x) \left[ \int |y - f(x)| P(y|x)dy \right] dx \)

when \( c=f(x) \) for a fixed \( x \)

\[ F(c) = \int |y-c| P(y|x)dy = \int_{-\infty}^{c} (c-y) P(y|x)dy + \int_{c}^{\infty} (y-c) P(y|x)dy \]

By applying Leibnitz rule

\[ F'(c) = \int_{-\infty}^{c} P(y|x)dy - \int_{c}^{\infty} P(y|x)dy = 0 \]

so \( \int_{-\infty}^{c} P(y|x)dy = \int_{c}^{\infty} P(y|x)dy \Rightarrow c \) is the conditional median

2 Problem 2

(a) \( f_1(x) \geq \min_x f_1(x) \)

\[ f_2(x) \geq \min_x f_2(x) \rightarrow \]

\[ f_1(x) + f_2(x) \geq \min_x f_1(x) + \min_x f_2(x) \rightarrow \]

\[ \min_x (f_1(x) + f_2(x)) \geq \min_x f_1(x) + \min_x f_2(x) \]

(b) \( g(a) + g(b) = \min_{R_1} \text{Risk}(a, R_1) + \min_{R_1} \text{Risk}(b, R_1) \leq \]

\[ \min_{R_1} [\text{Risk}(a, R_1) + \text{Risk}(b, R_1)] \text{[by (a)]} \]

\[ = 2 \min_{R_1} \{C_{22} + (C_{12} - C_{22}) \int_{R_1} P(x|H_2)dx + \frac{a + b}{2} [(C_{11} - C_{22})+} \]
\[(C_{21} - C_{11}) \int_{R_2} P(x|H_1)dx - (C_{12} - C_{22}) \int_{R_1} P(x|H_2)dx\]
\[= 2g\left(\frac{a+b}{2}\right)\]
\[\Rightarrow g(a) + g(b) \leq g\left(\frac{a+b}{2}\right)\]
\[\Rightarrow g(P_1) \text{ is concave down}\]

(c) If \(\text{Risk}(P_1, R_1)\) intersects \(g(P_1)\) at any \(P_1=a, P_1=b\), then because \(g(P_1)\) is concave down (assuming there is no flat part in \(g(P_1)\) curve)
\[g\left(\frac{a+b}{2}\right) > \text{Risk}\left(\frac{a+b}{2}, R_1\right)\]
That is impossible because \(g(P_1)\) is the best minimum risk at all points \(P_1\).
Also note that \(g(a) = \text{Risk}(a, R_1)\) so \(\text{Risk}(P_1, R_1)\) intersects \(g(P_1)\) exactly at one point, so the line \(\text{Risk}(R_1, R_1^*)\) is tangent to \(g(P_1)\).

3 Problem 3

Likelihood ratio test
\[P(x|H_1) > \frac{P_0}{P_1} = \frac{1/3}{2/3} = \frac{1}{2}\]
\[\frac{1}{\sqrt{2\pi}\delta_1} \exp\left(-\frac{(x-u_1)^2}{2\delta_1^2}\right) = 2\]
\[\frac{(x-u_1)^2}{\delta_1^2} - \frac{(x-u_2)^2}{\delta_2^2} = 2\ln\frac{2\delta_1}{\delta_2}\]
The classifier is the solution of \(x\) from this equation.

4 Problem 4

(a) \(\text{Risk}(T, P_1) = \int_{-\infty}^{T} P(x|H_2)dx + P_1 \left[ \int_{T}^{\infty} P(x|H_1)dx - \int_{-\infty}^{T} P(x|H_2)dx \right]\]
\[= \frac{1}{\sqrt{2\pi}\delta_2} \int_{-\infty}^{T} \exp\left(-\frac{(x-u_2)^2}{2\delta_2^2}\right)dx + \frac{P_0}{P_1} \left[ \frac{1}{\delta_1} \int_{T}^{\infty} \exp\left(-\frac{(x-u_1)^2}{2\delta_1^2}\right)dx - \frac{1}{\delta_2} \int_{-\infty}^{T} \exp\left(-\frac{(x-u_2)^2}{2\delta_2^2}\right)dx \right]\]
\[b(T) = 0 \Rightarrow \frac{1}{\delta_1} \int_{T}^{\infty} \exp\left(-\frac{(x-u_1)^2}{2\delta_1^2}\right)dx = \frac{1}{\delta_2} \int_{-\infty}^{T} \exp\left(-\frac{(x-u_2)^2}{2\delta_2^2}\right)dx\]
\[\Rightarrow \int_{T}^{\infty} \exp\left(-\frac{x^2}{2}\right)dx = \int_{-\infty}^{T} \exp\left(-\frac{x^2}{2}\right)dx\]

from the symmetry of \(\exp\left(-\frac{x^2}{2}\right)\)
\[\frac{T-u_1}{\delta_1} = -\frac{T-u_2}{\delta_2} \Rightarrow T = \frac{\frac{u_1}{\delta_1} + \frac{u_2}{\delta_2}}{\frac{1}{\delta_1} + \frac{1}{\delta_2}}\]

(b) minimax risk \(a(T) = \int_{-\infty}^{T} P(x|H_2)dx = \frac{1}{\sqrt{2\pi}\delta_2} \int_{-\infty}^{T} \exp\left(-\frac{(x-u_2)^2}{2\delta_2^2}\right)dx\]
\[= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{T} \exp\left(-\frac{x^2}{2}\right)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{u_2-u_1}{\delta_1+\delta_2}} e^{-\frac{t^2}{2}}dt = \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{u_2-u_1}{\delta_1+\delta_2}} e^{-\frac{t^2}{2}}dt = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_{\frac{u_2-u_1}{\delta_1+\delta_2}}^{\infty} e^{-t^2}dt =\]
\[
\frac{1}{2} - \frac{1}{2} \text{erf} \left( \frac{u - y_1}{\sqrt{2} \sigma_1} \right)
\]

(c) \text{Risk}(P_1, T) = a(T) + P_1 b(T) \text{ at the minimax point } \frac{d\text{Risk}(P_1, T)}{dT} = 0 \Rightarrow

\[
\frac{d}{dT} \left[ \int_{-\infty}^{T} P(x|H_2)dx + P_1 \left( \int_{T}^{\infty} P(x|H_1)dx - \int_{-\infty}^{T} P(x|H_2)dx \right) \right] = P(T|H_2) - P_1 P(T|H_1) - P_1 P(T|H_2) = 0 \Rightarrow P_1 = \frac{P(T|H_2)}{P(T|H_1) + P(T|H_2)}