Homework 2

Solution 1. We can write

$$\begin{aligned}
\arg\min_f \mathbb{E}l_{\alpha,\beta}(f(x),y) &= \arg\min_f \mathbb{E}_{X,Y}[\alpha\mathbf{1}\{f(X)=1, Y=0\} + \beta\mathbf{1}\{f(X)=0, Y=1\}] \\
&= \arg\min_f \mathbb{E}_X[\mathbb{E}_{Y|X}[\alpha\mathbf{1}\{f(X)=1, Y=0\} + \beta\mathbf{1}\{f(X)=0, Y=1\}]] \\
&= \arg\min_f \mathbb{E}_X[\int_y \alpha\mathbf{1}\{f(X)=1, y=0\} + \beta\mathbf{1}\{f(X)=0, Y=1\}dP(y|x)] \\
&= \arg\min_f \int_x [\alpha\mathbf{1}\{f(x)=1\}P(y=0|x) + \beta\mathbf{1}\{f(x)=0\}P(y=1|x)]dP(x)
\end{aligned}$$

We may minimize the integrand at each x by taking:

$$f(x) = \begin{cases} 1 & \beta P(y=1|x) \geq \alpha P(y=0|x) \\ 0 & \alpha P(y=0|x) > \beta P(y=1|x). \end{cases}$$

Solution 2. 1. The data likelihood of the normal distribution can be write as:

$$P(x_1,\ldots,x_n|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x_i-\mu)^2}{2\sigma^2})$$

And maximizing the log likelihood:

$$\log P(x_1,\ldots,x_n|\mu) = \sum_{i=1}^{n} -\frac{(x_i-\mu)^2}{2\sigma^2} + C$$

Here, C is a constance relevant with the known variance. So the μ for the maximum of log likelihood will be MLE estimator:

$$\mu_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. The prior for μ is:

$$P(\mu) = \frac{1}{\sqrt{2\pi}\beta}\exp(-\frac{(\mu-\nu)^2}{2\beta^2})$$

So, the posterior for μ will be:

$$
\begin{aligned}
P(\mu|Data) &= C' \cdot P(\mu) \cdot P(Data|\mu) \\
&= \frac{1}{\sqrt{2\pi}\beta} \exp(-\frac{(\mu-\nu)^2}{2\beta^2}) \cdot \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x_i-\mu)^2}{2\sigma^2}) \\
C' &= \frac{1}{P(Data)}
\end{aligned}
$$

And maximizing the log likelihood:

$$\log P(\mu|Data) = -\frac{(\mu-\nu)^2}{2\beta^2} + \sum_{i=1}^{n} -\frac{(x_i-\mu)^2}{2\sigma^2} + C''$$

3. Set the derivative of this formula to zero, and we will get:

$$\mu_{MAP} = \frac{\frac{\nu}{\beta^2} + \frac{\sum_{i=1}^{n} x_i}{\sigma^2}}{\frac{1}{\beta^2} + \frac{n}{\sigma^2}}$$

If the number of sample N goes to infinity, these two will be the same, both $\frac{1}{n}\sum_{i=1}^{n} x_i$.

Solution 3. 1. No Smoothing:

$P(Class = X) = 4/6 = 2/3; P(Class = Y) = 1/3$

$P(A1 = 0|Class = X) = 2/4 = 1/2; P(A1 = 1|Class = X) = 1/4; P(A1 = 2|Class = X) = 1/4$

$P(A2 = 0|Class = X) = 0; P(A2 = 1|Class = X) = 3/4; P(A2 = 2|Class = X) = 1/4$

$P(A1 = 0|Class = Y) = 0; P(A1 = 1|Class = Y) = 1/2; P(A1 = 2|Class = Y) = 1/2$

$P(A2 = 0|Class = Y) = 1/2; P(A2 = 1|Class = Y) = 0; P(A2 = 2|Class = Y) = 1/2$

And for the new example: $P(A1, A2, Class = X) = 0, P(A1, A2, Class = Y) = 1/4$, so its class will be predicted as Y.

2. With add-one smoothing:

$P(Class = X) = 5/8; P(Class = Y) = 3/8;$

$P(A1 = 0|Class = X) = 3/7; P(A1 = 1|Class = X) = 2/7; P(A1 = 2|Class = X) = 2/7$

$P(A2 = 0|Class = X) = 1/7; P(A2 = 1|Class = X) = 4/7; P(A2 = 2|Class = X) = 2/7$

$P(A1 = 0|Class = Y) = 1/5; P(A1 = 1|Class = Y) = 2/5; P(A1 = 2|Class = Y) = 2/5$

$P(A2 = 0|Class = Y) = 2/5; P(A2 = 1|Class = Y) = 1/5; P(A2 = 2|Class = Y) = 2/5$

And for the new example: $P(A1, A2, Class = X) = 20/392, P(A1, A2, Class = Y) = 12/200$, so its class will still be predicted as Y.