IE 7374 ST: Machine Learning in Engineering
Spring, 2021, Mid-Term
Solution Sheet

Rohit Bokade
NUID: 001280767

March 11, 2021

1. (1) **True.** For any continuous random variable x with a probability distribution function $p(x)$ the probability is bounded as $0 \leq p(x) \geq 1$ as $\int_{-\infty}^{\infty} p(x)dx = 1$

   (2) **False**. If X and Y are independent, then $\mathbb{E}[XY^2] = \mathbb{E}[X]\mathbb{E}[Y^2]$

   (3) **True**. True error is a hypothetical error value on the entire population. Hence, as $n \to \infty$ the training error will converge to the true error.

   (4) **False**. The linear regression is likely to have high bias and low variance whereas the polynomial regression with degree 3 is likely to have a low bias and high variance.

   (5) **False**. The empirical risk of f_1 will be greater than f_2.

   (6) **True**. Overfitting can occur due to a high approximation error as a result of overly simplistic hypothesis space or small quantity of data.

   (7) **False**. The approximation error will always be deterministic.

   (8) **True**. Linear regression has the lease variance.

   (9) **False**. The log likelihood of logistic regression is a convex function and will not have multiple local optima.

   (10) **False**. The correspondence implies that the posterior Gaussian Naive Bayes can be written in the same form as Logistic Regression.

2. (a)

$$p(y|x_1 = 0) = \frac{p(x_1 = 0|y)p(y)}{p(x_1 = 0)}$$

We know that,

$$p(x_1 = 0|y = 1) = 1 - \theta_1 = 0.5$$
$$p(x_1 = 0|y = 2) = 1 - \theta_2 = 0.5$$
$$p(x_1 = 0|y = 3) = 1 - \theta_3 = 0.5$$
$$p(y = 1) = 0.5, p(y = 2) = p(y = 3) = 0.25$$
$$p(x_1 = 0) = \sum_{m=1}^{3} p(x_1 = 0)p(y = m) = 0.25 + 0.125 + 0.125 = 0.5$$
$$p(y = 1|x_1 = 0) = \frac{0.5 \cdot 0.5}{0.5} = 0.5$$

Similarly,

$$p(y = 2|x_1 = 0) = \frac{0.5 \cdot 0.25}{0.25} = 0.25$$
$$p(y = 3|x_1 = 0) = \frac{0.5 \cdot 0.25}{0.25} = 0.25$$

(b)

$$p(y|x_2 = 0) = \frac{p(x_2 = 0|y)p(y)}{p(x_2 = 0)}$$
$$p(x_2 = 0|y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{1}{2\sigma^2}(x-\mu)^2} = \begin{bmatrix} 0.242 \\ 0.399 \\ 0.242 \end{bmatrix}$$
$$p(x_2 = 0) = \sum_{m=1}^{3} p(x_2 = 0)p(y = m) = 0.2815$$

Therefore,

$$p(y = 1|x_2 = 0) = \frac{0.242 \cdot 0.5}{0.2815} = 0.43$$
$$p(y = 2|x_2 = 0) = \frac{0.399 \cdot 0.5}{0.2815} = 0.354$$
$$p(y = 3|x_2 = 0) = \frac{0.242 \cdot 0.5}{0.2815} = 0.215$$

(c) $p(y = 2|x_1) = p(y = 3|x_1)$. Probability of classifying y as 2 or 3, given x_1, are equal .

3. (a)

$$p(z_{ik} = 1|p(w_i) = p(w_i)$$
$$p(z_{ik} = 0|p(w_i) = 1 - p(w_i)$$
$$p(z_{ik}|w_i) = p(w_i)^{z_{ik}}(1 - p(w_i))^{z}_{ik}$$

Using the law of independence, we have,

$$
\begin{aligned}
p(z_{ik}|w_i) &= p(w_i)^{z_{ik}}(1-p(w_i))^{z}_{ik} \\
&= \prod_{k=1}^{n} p(z_{ik}|p(w_i)) \\
&= \prod_{k=1}^{n} (p(w_i)^{z}_{ik}(1-p(w_i))^{(1-z_{ik})}
\end{aligned}
$$

(b) From the likelihood function derived above we can take log on both sides and rewrite the equation as

$$
\begin{aligned}
\ln p(w_i) &= \ln(\prod_{k=1}^{n} (p(w_i)^{z_{ik}}(1-p(w_i))^{(1-z_{ik})}) \\
&= \sum_{k=1}^{n} z_{ik}\ln(p(w_i) + (1-z_{ik})\ln(1-p(w_i))
\end{aligned}
$$

Taking derivative with respect to $p(w_i)$ and equating it to zero, we get

$$
\begin{aligned}
\frac{\partial \ln p(w_i)}{\partial p(w_i)} &= \frac{1}{p(w_i)}\sum_{i=1}^{n} z_{ik} - \frac{1}{1-p(w_i)}\sum_{k=1}^{n}(1-z_{ik}) \\
& \frac{1}{p(w_i)}\sum_{i=1}^{n} z_{ik} - \frac{1}{1-p(w_i)}\sum_{k=1}^{n}(1-z_{ik}) = 0 \\
\therefore (1-p(w_i))\sum_{i=1}^{n} z_{ik} &= p(w_i)\sum_{k=1}^{n}(1-z_{ik}) \\
\sum_{i=1}^{n} z_{ik} &= p(w_i)\sum_{i=1}^{n} z_{ik} + np(w_i) - p(w_i)\sum_{i=1}^{n} z_{ik} \\
\therefore p(w_i) &= \frac{1}{n}\sum_{i=1}^{n} z_{ik}
\end{aligned}
$$

4.

$$
p(y=1|x) = \frac{p(x|y=1)p(y=1)}{p(y=1)p(x|y=1) + p(y=2)p(x|y=2) + p(y=3)p(x|y=3)}
$$

5. Using Bayes' theorem we can write

$$
p(\theta|x<3) = \frac{p(x<3|\theta)p(\theta)}{p(x<3)}
$$

Prior can be written as the pdf of Beta distribution $= Beta(\theta, 1, 1) = \frac{1}{Beta(1,1)}\theta(1-\theta)^{1-1} = Uniform(0,1)$

Further, the likelihood can be given as

$$
\begin{aligned}
p(x < 3|\theta) &= \sum_{i=1}^{3} p(x = i|\theta) \\
&= \binom{5}{0} \theta^0 (1-\theta)^5 + \binom{5}{1} \theta^2 (1-\theta)^4 + \binom{5}{2} \theta^2 (1-\theta)^3 \\
&= \sum_{k=0}^{2} Binomial(k|\theta, 5)
\end{aligned}
$$

Therefor, the posterior can then be approximated as

$$
\begin{aligned}
p(\theta|x < 3) &\approx p(x < 3|\theta) p(\theta) \\
&= (\sum_{k=0}^{2} Binomial(k|\theta, 5)) \cdot Uniform(0, 1)
\end{aligned}
$$