Machine Learning HW4 - Theory

1.

a) **Question:** To ensure that $\mathbb{P}[X = i]$ is a valid probability mass function, what constraint should we put on $\theta = [\theta_1, \theta_2]$?

Answer: The sum of all of the probability components has to total 1.

$$\mathbb{P}[X = 1] + \mathbb{P}[X = 2] + \mathbb{P}[x = 3] = 1 \iff \theta_1 + 2\theta_1 + \theta_2 = 1 \iff 3\theta_1 + \theta_2 = 1$$

b) **Question:** Write down the joint probability of the data sequence

$$\mathbb{P}[D|\theta] = \mathbb{P}\left[\left\{x^{(1)}, .., x^{(n)}\right\}|\theta\right]$$

and the log probability $\log \mathbb{P}[D|\theta]$.

Answer: Since $x^{(i)}$ are i.i.d. then

$$\begin{aligned}
\mathbb{P}[D|\theta] &= \prod_{i=1}^{n} \mathbb{P}[x^{(i)}|\theta] \\
&= \prod_{i=1}^{n} \mathbb{P}[x^{(i)}|\theta] \\
&= \prod_{i=1}^{n_1} \theta_1 \prod_{i=1}^{n_2} 2\theta_1 \prod_{i=1}^{n_3} \theta_2 \\
&= \theta_1^{n_1} (2\theta_1)^{n_2} \theta_2^{n_3} \iff \\
l(\theta) = \log \mathbb{P}[D|\theta] &= n_1 \log(\theta_1) + n_2 \log(2\theta_1) + n_3 \log(\theta_2) \\
&= n_1 \log(\theta_1) + n_2 \log(2\theta_1) + n_3 \log(1 - 3\theta_1)
\end{aligned}$$

c) **Question:** Calculate the maximum likelihood estimation $\hat{\theta}$ of θ based on the sequence D.

Answer

$$
\begin{aligned}
\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ l(\theta) &= n_1 \log(\theta_1) + n_2 \log(2\theta_1) + n_3 \log(1 - 3\theta_1) \iff \\
\nabla l(\theta) &= \frac{n_1}{\theta_1} + \frac{n_2}{2\theta_1} - \frac{n_3}{1 - 3\theta_1} = 0 \iff \\
0 &= \frac{(2n_1 + n_2)(1 - 3\theta_1) - 2n_3\theta_1}{2\theta_1(1 - 3\theta_1)} \iff \\
0 &= \frac{(2n_1 + n_2)(1 - 3\theta_1) - 2n_3\theta_1}{2\theta_1(1 - 3\theta_1)} \iff \text{(assuming } \theta_1 > 0, \theta_2 > 0) \\
0 &= 2n_1 + n_2 - 6n_1\theta_1 - 3n_2\theta_1 - 2n_3\theta_1 \iff \\
\hat{\theta}_1 &= \frac{2n_1 + n_2}{6n_1 + 3n_2 + 2n_3} \\
\hat{\theta}_2 &= \frac{2n_3}{6n_1 + 3n_2 + 2n_3}
\end{aligned}
$$

and it checks out that $\hat{\theta}_1 + 2\hat{\theta}_1 + \hat{\theta}_2 = 1$.

2. **Question:** Let $x^{(i)}$ be i.i.d. from a distribution whose density is

$$
f(x|\beta) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right), 0 \le x \le \infty
$$

Find the MLE of β.

Answer: Since $x^{(i)}$ are i.i.d. then

$$
\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} f(x^{(i)}|\beta) \iff \\
l(\beta) &= \log \prod_{i=1}^{n} f(x^{(i)}|\beta) \\
&= \sum_{i=1}^{n} \log f(x^{(i)}|\beta) \\
&= \sum_{i=1}^{n} \log\left(\frac{1}{\beta} \exp\left(-\frac{x^{(i)}}{\beta}\right)\right) \\
&= \sum_{i=1}^{n} \log \frac{1}{\beta} - \frac{x^{(i)}}{\beta} \\
&= n \log \frac{1}{\beta} - \frac{\sum_{i=1}^{n} x^{(i)}}{\beta}
\end{aligned}
$$

To find $\underset{\beta}{\operatorname{argmax}}\ l(\beta)$ let's look at the derivative

$$
\begin{aligned}
\nabla l(\beta) = n\frac{1}{\frac{1}{\beta}}\left(-\frac{1}{\beta^2}\right) + \frac{\sum_{i=1}^{n} x^{(i)}}{\beta^2} &= 0 \iff \\
-n\frac{1}{\beta} + \frac{\sum_{i=1}^{n} x^{(i)}}{\beta^2} &= 0 \iff \text{(assuming } \beta > 0) \\
\hat{\beta} &= \frac{\sum_{i=1}^{n} x^{(i)}}{\beta}
\end{aligned}
$$

3.

a) **Question:** Assume that you want to investigate the proportion (θ) of defective items manufactured at a production line. You take a random sample of 30 items and found 5 of them were defective. Assume the prior of θ is a uniform distribution on $[0, 1]$. Please compute the posterior of θ. It is sufficient to write down the posterior density function upto a normalization constant that does not depend on θ.

Answer: Knowing our sample is i.i.d.

$$
\begin{aligned}
\mathbb{P}[\theta] &= \int_0^1 x dx = 1 \text{(prior)} \\
\mathbb{P}[x^{(i)} = \text{defect}|\theta] &= \theta \\
\mathbb{P}[x^{(i)} = \text{good}|\theta] &= 1 - \theta \\
\mathbb{P}[D] &= \text{const} \\
\mathbb{P}[\theta|D] &= \frac{\mathbb{P}[D|\theta]\mathbb{P}[\theta]}{\mathbb{P}[D]} \\
&= \frac{\prod_{i=1}^{n} \mathbb{P}[x^{(i)}|\theta] \times 1}{\mathbb{P}[D]} \\
&= \frac{\theta^5(1-\theta)^{25}}{\mathbb{P}[D]}
\end{aligned}
$$

b) **Question:** Assume an observation $D := \{x^i\}$ is i.i.d. drawn from a Gaussian distribution $\mathcal{N}(\mu, 1)$, with an unknown mean μ and a variance of 1. Assume the prior distribution of μ is $\mathcal{N}(0, 1)$. Please derive the posterior distribution $p(\mu|D)$ of μ given data D.

Answer: Knowing our sample is i.i.d. then $\mathbb{P}[D|\mu] = \prod_{i=1}^{n} \mathbb{P}[x^{(i)}|\mu]$ and using the same notation for A and B for the coefficients of μ from the quadratic from lecture

$$
\begin{aligned}
\mathbb{P}[\mu] &= \mathcal{N}(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2}\right) \text{(prior)} \\
\mathbb{P}[x^{(i)}|\mu] &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x^{(i)}-\mu)^2}{2}\right) \\
\mathbb{P}[D] &= \text{const} \\
\mathbb{P}[\mu|D] &= \frac{\mathbb{P}[D|\mu]\mathbb{P}[\mu]}{\mathbb{P}[D]} \\
&= \frac{\prod_{i=1}^{n} \mathbb{P}[x^{(i)}|\mu] \times \mathbb{P}[\mu]}{\mathbb{P}[D]} \\
&= \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x^{(i)}-\mu)^2}{2}\right) \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2}\right)}{\mathbb{P}[D]} \\
&= \frac{\frac{1}{(2\pi)^{\frac{n+1}{2}}} \exp\left(\sum_{i=1}^{n} -\frac{(x^{(i)}-\mu)^2}{2} - \frac{\mu^2}{2}\right)}{\mathbb{P}[D]} \iff \\
A &= \sum_{i=1}^{n} 1 + 1 = n+1 \\
B &= \sum_{i=1}^{n} x^{(i)} \\
\mu_p &= \frac{\sum_{i=1}^{n} x^{(i)}}{n+1} \iff \\
\mathbb{P}[\mu|D] &\propto \mathcal{N}(\mu_p, 1)
\end{aligned}
$$