Bayesian Machine Learning: Assignment #1

Prof. Juho Lee

Hyunsu Kim

Problem 1

(a) Using Integration by parts,

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, \mathrm{d}x = \int_{-\infty}^{\theta} x \cdot 0\, \mathrm{d}x + \int_{\theta}^{\infty} x e^{\theta - x}\, \mathrm{d}x = 1 + \theta$$

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \int_{\theta}^{\infty} x^2 e^{\theta - x}\, \mathrm{d}x - (1+\theta)^2 = \theta^2 + 2(1+\theta) - (1+\theta)^2 = 1$$

(b) To show unbiasedness, compare θ with the exepectation. Note X_i's are i.i.d., and $\mathrm{E}[X_i] = 1 + \theta$ in (a):

$$
\begin{aligned}
\mathrm{E}[\hat{\theta}_n] &= \mathrm{E}[\frac{1}{n}\sum_{i=1}^{n}(X_i - 1)] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i - 1] \\
&= \frac{1}{n}\sum_{i=1}^{n}(\mathrm{E}[X_i] - 1) \\
&= \frac{1}{n}\sum_{i=1}^{n}((1+\theta) - 1) \\
&= \frac{1}{n}\sum_{i=1}^{n}\theta \\
&= \frac{1}{n}(n\theta) \\
&= \theta
\end{aligned}
$$

(c) Let $Y_i = X_i - 1$. Then, $\mathrm{E}[Y_i] = \theta$ and $\mathrm{Var}[Y_i] = 1$. Now, according to (b) and Central Limit Theorem,

$$\frac{\hat{\theta}_n - \theta}{1/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

holds. Then, $100(1 - \alpha)\%$ confidence interval of θ is $[\hat{\theta}_n - \frac{\Phi^{-1}(1 - \frac{\alpha}{2})}{\sqrt{n}}, \hat{\theta}_n - \frac{\Phi^{-1}(\frac{\alpha}{2})}{\sqrt{n}}]$ where Φ is Cumulative Density Function (CDF) of unit normal distribution.

(d) According to given fact and that unit normal distribution is an even function, $\Phi^{-1}(0.025) = -1.96$ and $\Phi^{-1}(0.975) = 1.96$. Thus, 95% confidence interval of θ is $[11.33 - \frac{1.96}{\sqrt{3}}, 11.33 + \frac{1.96}{\sqrt{3}}] = [10.20, 12.46]$. It is weird that the observed data 10.0 is contradictory for any θ in the obtained confidence interval according to the PDF of X.

Problem 2

1. The proof is straightforward:

$$
\begin{aligned}
\mu(B) &= \mu(B \cap (A \cup A^c)) \\
&= \mu((B \cap A) \cup (B \cap A^c)) && \because \text{distributive law} \\
&= \mu(B \cap A) + \mu(B \cap A^c) && \because \text{countable additivity of } \mu \\
&= \mu(A) + \mu(B \cap A^c) && \because A \subset B \implies B \cap A = A \\
&\geq \mu(A) && \because \mu(B \cap A^c) \geq 0 : \text{nonnegativity of } \mu
\end{aligned}
$$

                                        2

2. Use induction. (Base case) We first show it holds for $n = 1$:

$$\mu(\bigcup_{i=1}^{1} A_i) = \mu(A_1)$$

$$= \sum_{i=1}^{1} \mu(A_i)$$

$$\leq \sum_{i=1}^{1} \mu(A_i)$$

(Inductive case) Assume it holds for $n = k$. Want to show it holds for $n = k + 1$:

$$\mu(\bigcup_{i=1}^{k+1} A_i) = \mu(\bigcup_{i=1}^{k} A_i \cup A_{k+1})$$

$$= \mu(\bigcup_{i=1}^{k} A_i \cup (A_{k+1} - \bigcup_{i=1}^{k} A_i)) \qquad \because \text{simple Venn diagram argument}$$

$$= \mu(\bigcup_{i=1}^{k} A_i) + \mu(A_{k+1} - \bigcup_{i=1}^{k} A_i) \qquad \because \text{countable additivity of } \mu$$

$$\leq \mu(\bigcup_{i=1}^{k} A_i) + \mu(A_{k+1}) \qquad \because (A_{k+1} - \bigcup_{i=1}^{k} A_i) \subseteq A_{k+1} \text{ and 1.}$$

$$\leq \sum_{i=1}^{k} \mu(A_i) + \mu(A_{k+1}) \qquad \because \text{inductive hypothesis}$$

$$= \sum_{i=1}^{k+1} \mu(A_i)$$

$\square$

Problem 3

To show the asserted convergence in probability, want to show

$$\lim_{n \to \infty} \mathbb{P}(|\overline{X}_n - \mathrm{E}[X_1]| > \epsilon) = 0$$

for any $\epsilon > 0$. Note $\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$ is a sample mean of i.i.d. random variables so that $\mathrm{E}[\overline{X}_n] = \mathrm{E}[X_1]$ and $\mathrm{Var}[\overline{X}_n] = \frac{\mathrm{Var}[X_1]^2}{n}$. Using Chebyshev's inequality for $k > 0$:

$$\mathbb{P}(|\overline{X}_n - \mathrm{E}[X_1]| \geq k \frac{\mathrm{Var}[X_1]}{\sqrt{n}}) \leq \frac{1}{k^2} \tag{1}$$

holds. Now, let $k = \frac{\sqrt{n}\epsilon}{\mathrm{Var}[X_1]}$. Then, (1) becomes as follows:

$$\mathbb{P}(|\overline{X}_n - \mathrm{E}[X_1]| \geq \epsilon) \leq \frac{Var[X_1]^2}{n\epsilon^2}$$

Then, as we take $n \to \infty$ on both sides of inequality:

$$\lim_{n \to \infty} \mathbb{P}(|\overline{X}_n - \mathrm{E}[X_1]| > \epsilon) \leq \lim_{n \to \infty} \frac{Var[X_1]^2}{n\epsilon^2} = 0$$

holds for any $\epsilon > 0$. $\square$

Problem 4

Give an appropriate positive constant c such that $f(n) \leq c \cdot g(n)$ for all $n > 1$.

1. $f(n) = n^2 + n + 1$, $g(n) = 2n^3$

2. $f(n) = n\sqrt{n} + n^2$, $g(n) = n^2$

3. $f(n) = n^2 - n + 1$, $g(n) = n^2/2$

Solution
We solve each solution algebraically to determine a possible constant c.

Part One

$$
\begin{aligned}
n^2 + n + 1 &= \\
&\leq n^2 + n^2 + n^2 \\
&= 3n^2 \\
&\leq c \cdot 2n^3
\end{aligned}
$$

Thus a valid c could be when $c = 2$.

Part Two

$$
\begin{aligned}
n^2 + n\sqrt{n} &= \\
&= n^2 + n^{3/2} \\
&\leq n^2 + n^{4/2} \\
&= n^2 + n^2 \\
&= 2n^2 \\
&\leq c \cdot n^2
\end{aligned}
$$

Thus a valid c is $c = 2$.

Part Three

$$
\begin{aligned}
n^2 - n + 1 &= \\
&\leq n^2 \\
&\leq c \cdot n^2/2
\end{aligned}
$$

Thus a valid c is $c = 2$.

Problem 5

Let $\Sigma = \{0,1\}$. Construct a DFA A that recognizes the language that consists of all binary numbers that can be divided by 5.

Let the state q_k indicate the remainder of k divided by 5. For example, the remainder of 2 would correlate to state q_2 because $7 \mod 5 = 2$.
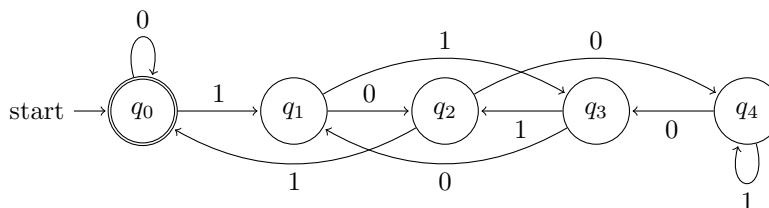

Figure 1: DFA, A, this is really beautiful, ya know?

Justification

Take a given binary number, x. Since there are only two inputs to our state machine, x can either become $x0$ or $x1$. When a 0 comes into the state machine, it is the same as taking the binary number and multiplying it by two. When a 1 comes into the machine, it is the same as multipying by two and adding one.

Using this knowledge, we can construct a transition table that tell us where to go:

	$x \mod 5 = 0$	$x \mod 5 = 1$	$x \mod 5 = 2$	$x \mod 5 = 3$	$x \mod 5 = 4$
$x0$	0	2	4	1	3
$x1$	1	3	0	2	4

Therefore on state q_0 or $(x \mod 5 = 0)$, a transition line should go to state q_0 for the input 0 and a line should go to state q_1 for input 1. Continuing this gives us the Figure 1.

Problem 6

Write part of **Quick-Sort**$(list, start, end)$

```
1: function QUICK-SORT(list, start, end)
2:      if start ≥ end then
3:          return
4:      end if
5:      mid ← PARTITION(list, start, end)
6:      QUICK-SORT(list, start, mid − 1)
7:      QUICK-SORT(list, mid + 1, end)
8: end function
```

Algorithm 1: Start of QuickSort

Problem 7

Suppose we would like to fit a straight line through the origin, i.e., $Y_i = \beta_1 x_i + e_i$ with $i = 1, \ldots, n$, $\mathrm{E}[e_i] = 0$, and $\mathrm{Var}[e_i] = \sigma_e^2$ and $\mathrm{Cov}[e_i, e_j] = 0, \forall i \neq j$.

Part A
Find the least squares esimator for $\hat{\beta}_1$ for the slope β_1.

Solution
To find the least squares estimator, we should minimize our Residual Sum of Squares, RSS:

$$RSS = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$
$$= \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 x_i\right)^2$$

By taking the partial derivative in respect to $\hat{\beta}_1$, we get:

$$\frac{\partial}{\partial \hat{\beta}_1}(RSS) = -2 \sum_{i=1}^{n} x_i(Y_i - \hat{\beta}_1 x_i) = 0$$

This gives us:

$$\sum_{i=1}^{n} x_i(Y_i - \hat{\beta}_1 x_i) = \sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} \hat{\beta}_1 x_i^2$$
$$= \sum_{i=1}^{n} x_i Y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2$$

Solving for $\hat{\beta}_1$ gives the final estimator for β_1:

$$\hat{\beta}_1 = \frac{\sum x_i Y_i}{\sum x_i^2}$$

       6

Part B

Calculate the bias and the variance for the estimated slope $\hat{\beta}_1$.

Solution

For the bias, we need to calculate the expected value $\mathrm{E}[\hat{\beta}_1]$:

$$
\begin{aligned}
\mathrm{E}[\hat{\beta}_1] &= \mathrm{E}\left[\frac{\sum x_i Y_i}{\sum x_i^2}\right] \\
&= \frac{\sum x_i \mathrm{E}[Y_i]}{\sum x_i^2} \\
&= \frac{\sum x_i (\beta_1 x_i)}{\sum x_i^2} \\
&= \frac{\sum x_i^2 \beta_1}{\sum x_i^2} \\
&= \beta_1 \frac{\sum x_i^2 \beta_1}{\sum x_i^2} \\
&= \beta_1
\end{aligned}
$$

Thus since our estimator's expected value is β_1, we can conclude that the bias of our estimator is 0.

For the variance:

$$
\begin{aligned}
\mathrm{Var}[\hat{\beta}_1] &= \mathrm{Var}\left[\frac{\sum x_i Y_i}{\sum x_i^2}\right] \\
&= \frac{\sum x_i^2}{\sum x_i^2 \sum x_i^2} \mathrm{Var}[Y_i] \\
&= \frac{\sum x_i^2}{\sum x_i^2 \sum x_i^2} \mathrm{Var}[Y_i] \\
&= \frac{1}{\sum x_i^2} \mathrm{Var}[Y_i] \\
&= \frac{1}{\sum x_i^2} \sigma^2 \\
&= \frac{\sigma^2}{\sum x_i^2}
\end{aligned}
$$

Problem 8

Prove a polynomial of degree k, $a_k n^k + a_{k-1} n^{k-1} + \ldots + a_1 n^1 + a_0 n^0$ is a member of $\Theta(n^k)$ where $a_k \ldots a_0$ are nonnegative constants.

Proof. To prove that $a_k n^k + a_{k-1} n^{k-1} + \ldots + a_1 n^1 + a_0 n^0$, we must show the following:

$$\exists c_1 \exists c_2 \forall n \geq n_0, \ c_1 \cdot g(n) \leq f(n) \leq c_2 \cdot g(n)$$

For the first inequality, it is easy to see that it holds because no matter what the constants are, $n^k \leq a_k n^k + a_{k-1} n^{k-1} + \ldots + a_1 n^1 + a_0 n^0$ even if $c_1 = 1$ and $n_0 = 1$. This is because $n^k \leq c_1 \cdot a_k n^k$ for any nonnegative constant, c_1 and a_k.

Taking the second inequality, we prove it in the following way. By summation, $\sum_{i=0}^{k} a_i$ will give us a new constant, A. By taking this value of A, we can then do the following:

$$
\begin{aligned}
a_k n^k + a_{k-1} n^{k-1} + \ldots + a_1 n^1 + a_0 n^0 = \\
\leq (a_k + a_{k-1} \ldots a_1 + a_0) \cdot n^k \\
= A \cdot n^k \\
\leq c_2 \cdot n^k
\end{aligned}
$$

where $n_0 = 1$ and $c_2 = A$. c_2 is just a constant. Thus the proof is complete. $\qquad\square$

Problem 18

Evaluate $\sum_{k=1}^{5} k^2$ and $\sum_{k=1}^{5} (k-1)^2$.

Problem 19

Find the derivative of $f(x) = x^4 + 3x^2 - 2$

Problem 6

Evaluate the integrals $\int_0^1 (1 - x^2)\, dx$ and $\int_1^{\infty} \frac{1}{x^2}\, dx$.

Problem 20

HI