Homework #0

Hyungseok Choi
Student No. : 2250166

Probability and Statistics

A.1

$$\begin{aligned} P(disease|positive) &= \frac{P(positive|disease) \cdot P(disease)}{P(positive)} \\ &= \frac{0.99 \cdot 0.0001}{0.0001 \cdot 0.99 + 0.9999 \cdot 0.01} \\ &= 0.9804 \\ \therefore 0.98\% \end{aligned}$$

A.2

a.

$$\begin{aligned} \mathbb{E}[Y|X=x] &= x \\ &= \textstyle\sum_y f(y|X=x) \\ &= \textstyle\sum_y y \cdot \frac{f(x,y)}{f(X=x)} \end{aligned}$$

$$\begin{aligned} x &= \textstyle\sum_y y \cdot \frac{f(x,y)}{f(X=x)} \\ xf(X=x) &= \textstyle\sum_y y \cdot f(x,y) \\ \sum_x xf(X=x) &= \textstyle\sum_x \sum_y y \cdot f(x,y) \\ \therefore \mathbb{E}(X) = \mathbb{E}(Y) \end{aligned}$$

$$\begin{aligned} \sum_x x^2 f(X=x) &= \textstyle\sum_x \sum_y xy \cdot f(x,y) \\ \therefore \mathbb{E}(X^2) = \mathbb{E}(XY) \end{aligned}$$

Therefore,

$$\begin{aligned} Cov(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\ &= \mathbb{E}[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] \\ &= \mathbb{E}[X-\mathbb{E}[X]]^2 \end{aligned}$$

b.

$$\begin{aligned} Cov(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\ &= \mathbb{E}[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= 0 \quad (\because \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \text{ for independent variables X, Y}) \end{aligned}$$

A.3

a.

$$
\begin{aligned}
H(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} h(x,y)\,dydx \\
&= \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{z-x} g(y)\,dydx \quad (independence) \\
&= \int_{-\infty}^{\infty} f(x) G(z-x)\,dy \\
&\left(\because [G(y)]_{-\infty}^{z-x} = G(z-x) - G(-\infty) = G(z-x)\right)
\end{aligned}
$$

$$
\begin{aligned}
h(z) &= \frac{dH(z)}{dz} \\
&= \int_{-\infty}^{\infty} f(x) \tfrac{dG(z-x)}{dz}\,dy \\
&= \int_{-\infty}^{\infty} f(x) g(z-x)\,dy
\end{aligned}
$$

b.

$$
\begin{aligned}
h(z) &= \int_{-\infty}^{\infty} f(x)g(z-x)\,dy \\
&= \int_{-\infty}^{0} f(x)g(z-x)\,dy \;+\; \int_{0}^{t} f(x)g(z-x)\,dy \;+\; \int_{t}^{1} f(x)g(z-x)\,dy \;+\; \int_{1}^{\infty} f(x)g(z-x)\,dy \\
&= \int_{0}^{t} 1\,dy \;+\; \int_{t}^{1} 1\,dy \\
&= 1
\end{aligned}
$$

A.4

When normalizing X, we get

$$Z \sim \mathcal{N}(0,\,1)\, st.\, Z = \frac{X-\mu}{\sigma}$$

Since Y = Z,

$$a \;=\; \frac{1}{\sigma} b \;=\; -\frac{\mu}{\sigma}$$

A.5

$$
\begin{aligned}
\mathbb{E}\left[\sqrt{n}\left(\hat{\mu_n}-\mu\right)\right] &= \sqrt{n}\left(\mathbb{E}(\hat{\mu_n})-\mu\right) \\
&= \sqrt{n}\left(\tfrac{1}{n}\mathbb{E}\left(\textstyle\sum_{i=1}^{n} X_i\right)-\mu\right) \\
&= \sqrt{n}\left(\tfrac{1}{n}\cdot(n\mu)-\mu\right) \\
&= \sqrt{n}\left(\mu-\mu\right) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
Var\left[\sqrt{n}\left(\hat{\mu_n}-\mu\right)\right] &= \mathbb{E}\left[\left(\sqrt{n}\left(\hat{\mu_n}-\mu\right)\right)^2\right]-\mathbb{E}\left[\sqrt{n}\left(\hat{\mu_n}-\mu\right)\right]^2 \\
&= n\cdot\mathbb{E}\left[\left(\hat{\mu_n}-\mu\right)^2\right]-0 \\
&= n\cdot\mathbb{E}\left[\left(\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} X_i\right)-\mu\right)^2\right] \\
&= \frac{1}{n}\cdot\mathbb{E}\left[\left(\textstyle\sum_{i=1}^{n}\left(X_i-\mu\right)\right)^2\right] \\
&= \frac{1}{n}\cdot\left[\mathbb{E}\left[\textstyle\sum_{i=1}^{n}\left(X_i-\mu\right)^2\right]+\mathbb{E}\left[\textstyle\sum\sum_{i\neq j}\left(X_i-\mu\right)\left(X_j-\mu\right)\right]\right] \\
&= \frac{1}{n}\cdot\sigma^2+0 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

$$
\because \mathbb{E}\left[\sum\sum_{i\neq j}\left(X_i-\mu\right)\left(X_j-\mu\right)\right]=\mathbb{E}\left[\sum\left(X_i-\mu\right)\cdot\sum\left(X_j-\mu\right)\right]
$$

$$
\text{and each term } \mathbb{E}\left[\left(X_i-\mu\right)\right]=\mathbb{E}\left[X_i\right]-\mu=0
$$

A.6

a.

$$\begin{aligned}\mathbb{E}\left[\hat{F_n}(x)\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} 1\left\{X_i \le x\right\}\right]\\ &= \frac{1}{n}\cdot\sum_{i=1}^{n}\left[1\left\{X_i \le x\right\}\right]\\ &= F(x)\end{aligned}$$

b.

$$\begin{aligned}Var\left(\hat{F_n}(x)\right) &= \mathbb{E}\left[\left(\hat{F_n}(x) - F(x)\right)^2\right]\\ &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} 1\left\{X_i \le x\right\} - F(x)\right)^2\right]\\ &= \frac{1}{n^2}\cdot\mathbb{E}\left[\left\{\sum_{i=1}^{n}\left(1\left\{X_i \le x\right\} - F(x)\right)\right\}^2\right]\\ &= \frac{1}{n^2}\cdot\left[\mathbb{E}\left[\sum_{i=1}^{n}\left(1\left\{X_i \le x\right\} - F(x)\right)^2\right] + \mathbb{E}\left[\sum\sum_{i\ne j}\left(1\left\{X_i \le x\right\}\right)\left(1\left\{X_j \le x\right\}\right)\right]\right]\\ &= \frac{1}{n^2}\cdot\mathbb{E}\left[\sum_{i=1}^{n}\left(1\left\{X_i \le x\right\} - F(x)\right)^2\right] + 0\,(\text{Same as A.5})\\ &= \frac{1}{n^2}\cdot\sum_{i=1}^{n}\left(\mathbb{E}\left[\left(1\left\{X_i \le x\right\}\right)^2 - 2\cdot\left(1\left\{X_i \le x\right\}\right)\cdot F(x) + F(x)^2\right]\right)\\ &= \frac{1}{n}\left(F(x) - 2\cdot F(x)^2 + F(x)^2\right)\\ &= \frac{F(x)\left(1 - F(x)\right)}{n}\end{aligned}$$

$$\left(\because \mathbb{E}\left(1\left\{X_i \le x\right\}\right)^2 = \mathbb{E}\left(1\left\{X_i \le x\right\}\right)\ \text{Since a squared of 1 is still 1}\ \right)$$

c. F(x) is CDF, $0 <= F(x) <= 1$.

$$Var\left(\hat{F_n}(x)\right)\ \text{is function of F(x). Let F(x) = t}$$

$$\frac{d}{dt}\left(\frac{t\,(1-t)}{n}\right) = \frac{1}{n}(1-2t)$$

$$\therefore \frac{F(x)\left(1 - F(x)\right)}{n} \le \frac{1}{4n}\ \text{When F(x)} = \frac{1}{2}$$

Linear Algebra and Vector Calculus

A.7 (Rank) Let $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$.

a.

$$\text{For A, } \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow{R3=R3-R2} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 3 \\ 0 & 1 & -1 \end{bmatrix} \xrightarrow{R2=R2-R1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & -2 \\ 0 & 1 & -1 \end{bmatrix} \xrightarrow[R3=R3-R2]{R2=R2/2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

Rank A = 2

$$\text{For B, } \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow{R3=R3-R2} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{R2=R1-R2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 2 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow[R3=R3-R2]{R2=R2/2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Rank B = 2

b.

$$\text{Basis for Matrix A } : C1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, C2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

$$\text{Basis for Matrix B } : C1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, C2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

A.8 (Linear equations) Let $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} -2 & -2 & -4 \end{bmatrix}^T$, and $c = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$.

a.

$$Ac = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \cdot 1 + 2 \cdot 1 + 4 \cdot 1 \\ 2 \cdot 1 + 4 \cdot 1 + 2 \cdot 1 \\ 3 \cdot 1 + 3 \cdot 1 + 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$$

b.

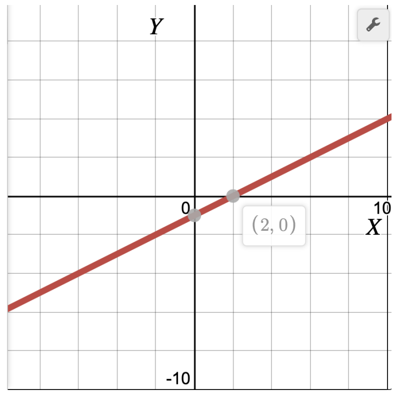
$$\left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 2 & 4 & 2 & -2 \\ 3 & 3 & 1 & -4 \end{array}\right] \xrightarrow{R2=R2/2} \left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 1 & 2 & 1 & -1 \\ 3 & 3 & 1 & -4 \end{array}\right] \xrightarrow{R3=R3-3R2} \left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 1 & 2 & 1 & -1 \\ 0 & -3 & -2 & -1 \end{array}\right] \xrightarrow{R3=R3+\frac{3}{2}R1} \left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 4 & -4 \end{array}\right]$$

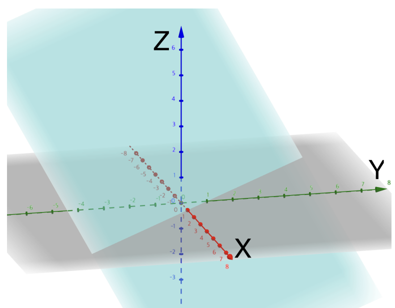
$$x_3 = -1 \text{ From } 4x_3 = -4$$

$$x_2 = 1 \text{ From } 2x_2 - 4 = -2$$

$$x_1 = -2 \text{ From } x_1 + 2 - 1 = -1$$

A.9 (Hyperplanes) Assume w is an n-dimensional vector and b is a scalar. A hyperplane in $\mathbb{R}^n$ is the set $\{x : x \in \mathbb{R}^n, \text{ s.t. } w^T x + b = 0\}$.

a. 

b. 

c. Let x be an arbitrary point on the hyperplane $w^T x + b = 0$. Then, the vector from x_0 to the plane is $x_0 - x$.
Since the dot product of two vectors $w^T(x_0 - x)$ represents $|w||(x_0 - x)|\cos\theta$, we can get the distance between the dot$(= x_0)$ and the plane by diving both sides by $|w|$.

$$
\begin{aligned}
d &= \frac{|w^T(x_0 - x)|}{|w|} \\
&= \frac{|w^T x_0 - w^T x|}{|w|} \\
&= \frac{|w^T x_0 + b|}{|w|}
\end{aligned}
$$

Then,

$$
\begin{aligned}
d^2 &= \left(\frac{|w^T x_0 + b|}{|w|}\right)^2 \\
&= \frac{(|w^T x_0 + b|)^2}{|w|^2} \\
&= \frac{(|w^T x_0 + b|)^2}{w^T w}
\end{aligned}
$$

A.10

a.

$$f(x,y) = \sum_{j=1}^{n}\sum_{i=1}^{n} A_{ij}x_ix_j + \sum_{j=1}^{n}\sum_{i=1}^{n} B_{ij}x_iy_j + c$$

b. Let P $= x^TAx$, Q $= y^TBx$

$$\begin{aligned}
\frac{\partial P}{\partial x_k} &= \frac{\partial}{\partial x_k}\textstyle\sum_{j=1}^{n}\sum_{i=1}^{n} A_{ij}x_ix_j \\
&= \frac{\partial}{\partial x_k}\left(\textstyle\sum_{i\neq k}\sum_{j=1}^{n} a_{ij}x_ix_j\right) + \frac{\partial}{\partial x_k}\left(\textstyle\sum_{j=1}^{n} a_{kj}a_ka_j\right) \\
&= \textstyle\sum_{i\neq k}\left(\frac{\partial}{\partial x_k}\left(\sum_{j\neq k} a_{ij}x_ix_j\right) + \frac{\partial}{\partial x_k}(a_{ik}x_ix_k)\right) + \sum_{j\neq k}\frac{\partial}{\partial x_k}(a_{kj}x_kx_j) + \frac{\partial}{\partial x_k}\left(a_{kk}x_k^2\right) \\
&= 0 + \textstyle\sum_{i\neq k} a_{ij}x_i + \sum_{j\neq k} a_{kj}x_j + 2a_{kk}x_k \\
&= \textstyle\sum_{i=1}^{n} a_{ik}x_i + \sum_{j=1}^{n} a_{kj}x_j
\end{aligned}$$

which means, k^{th} gradiant component of P is equal to the sum of k^{th} component of x^TA and k^{th} component of Ax which is equal to k^{th} component of x^TA^T By generalizing this, we get

$$\begin{aligned}
\nabla_x P &= x^TA + x^TA^T \\
&= x^T(A + A^T)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial Q}{\partial x_k} &= \frac{\partial}{\partial x_k}\textstyle\sum_{j=1}^{n}\sum_{i=1}^{n} B_{ij}x_iy_j \\
&= \frac{\partial}{\partial x_k}\left(\textstyle\sum_{i\neq k}\sum_{j=1}^{n} B_{ij}x_iy_j\right) + \frac{\partial}{\partial x_k}\left(\textstyle\sum_{j=1}^{n} B_{kj}x_ky_j\right) \\
&= \textstyle\sum_{j=1}^{n} B_{kj}y_j
\end{aligned}$$

Same as above, it means, k^{th} component of By which is equal to k^{th} component of y^TB^T By generalizing this, we get

$$\nabla_x Q = y^TB^T$$

$$\therefore \nabla_x f(x,y) = x^T(A + A^T) + y^TB^T$$

c. Let P $= x^TAx$, Q $= y^TBx$. Then $\nabla_y P = 0$.

$$\begin{aligned}
\frac{\partial Q}{\partial y_k} &= \frac{\partial}{\partial y_k}\textstyle\sum_{j=1}^{n}\sum_{i=1}^{n} B_{ij}x_iy_j \\
&= \frac{\partial}{\partial y_k}\left(\textstyle\sum_{j\neq k}\sum_{i=1}^{n} B_{ij}x_iy_j\right) + \frac{\partial}{\partial y_k}(\textstyle\sum_{i=1}^{n} B_{ik}x_iy_k) \\
&= \textstyle\sum_{i=1}^{n} B_{ik}x_i
\end{aligned}$$

Same as above, it means, k^{th} component of x^TB. By generalizing this, we get

$$\nabla_y Q = x^TB$$

$$\therefore \nabla_y f(x,y) = x^TB$$

Programming

A.11

a.

```
[4] import numpy as np
    import matplotlib.pyplot as plot
```

```
mat_A = np.matrix([
                   [0, 2, 4],
                   [2, 4, 2],
                   [3, 3, 1]
])

vec_bt = np.matrix([-2, -2, -4])
vec_b = np.transpose(vec_bt)

vec_ct = np.matrix([1, 1, 1])
vec_c = np.transpose(vec_ct)

print(np.transpose(mat_A))
print(np.multiply(np.transpose(mat_A), vec_b))
print(np.multiply(mat_A, vec_c))
```

```
[[0 2 3]
 [2 4 3]
 [4 2 1]]
[[  0  -4  -6]
 [ -4  -8  -6]
 [-16  -8  -4]]
[[0 2 4]
 [2 4 2]
 [3 3 1]]
```

b.

A.12

a.

```
n = 40000
Z = np.random.randn(n)

klist = [1,8,64,512]
for k in klist:
  sns.ecdfplot(data=np.sum(np.sign(np.random.randn(n, k))*np.sqrt(1./k), axis=1), label=str(k))

p = sns.ecdfplot(data=sorted(Z), label="Gaussian")
p.set_xlim(-3, 3)
p.set_xlabel("Observations", fontsize = 10)
p.set_ylabel("Probability", fontsize = 10)
plt.grid(True)
plt.legend(framealpha=0)
plt.show()
```

b.