Machine Learning

Assignment 1

1. Variance of sum:

$$var[X+Y] = var[X] + var[Y] + 2\,cov[X,Y]$$

we have:

(1) $var[X+Y] = E[(X+Y)^2] - E^2[X+Y]$

(2) $var[X] + var[Y] = E[X^2] - E^2[X] + E[Y^2] - E^2[Y]$

(3) $2cov[X,Y] = 2.E[(X-E[X])(Y-E[Y])]$

$= 2.E[XY - Y.E[X] - X.E[Y] + E[X].E[Y]]$

$= 2.E[XY] - 2E[Y.E[X]] - 2E[X.E[Y]] + 2E[E[X].E[Y]]$

$= 2.E[XY] - 2.E[X].E[Y] - 2.E[Y].E[X] + 2.E[X].E[Y]$

$= 2.E[XY] - 2.E[X].E[Y]$

And $(2)+(3) = E[X^2] - E^2[X] + E[Y^2] - E^2[Y] + 2E[XY] - 2E[X]E[Y]$

$= E[X^2] + 2E[XY] + E[Y^2] - (E^2[X] + 2E[X].E[Y] + E^2[Y])$

$= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2$

$= E[(X+Y)^2] - (E[X+Y])^2 = (1)$

So we can say that

$(1) = (2) + (3)$

2. Independence:

X, Y independent $\Rightarrow E[X+Y] = E[X] + E[Y]$

we have:

$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y).f_{X,Y}(x,y)\,dx.dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y).f_X(x).f_Y(y).dx.dy$$

// X, Y independent $\Rightarrow f_{XY}(x,y) = f_X(x).f_Y(y)$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x . f_X(x) . f_Y(y) . dx . dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y . f_X(x) . f_Y(y) dx . dy$$

$$= \int_{-\infty}^{\infty} x . f_X(x) . dx + \int_{-\infty}^{\infty} y f_Y(y) . dy$$

$$= E[X] + E[Y]$$

4. Mean and variance

$$E[Y] = E[a.X+b] = a E[X] + b = a . \mu + b$$

$$\text{var}[Y] = \text{var}[a.X+b] = E[(aX+b)^2] - E^2[aX+b]$$

$$= E[a^2X^2 + 2ab.X + b^2] - (a E[X] + b)^2$$

$$= a^2 . E[X^2] + 2ab E[X] + b^2 - (a^2 E^2[X] + 2ab E[X] + b^2)$$

$$= a^2 E[X^2] - a^2 E^2[X]$$

$$= a^2 (E[X^2] - E^2[X])$$

$$= a^2 . \text{var}[X] = a^2 . \sigma^2$$

6. Derivatives

$$f(x) = e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$\Rightarrow f'(x) = -\frac{1}{2\sigma^2} . 2(x-\mu) . 1 . e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$= -\frac{(x-\mu)}{\sigma^2} . e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

5. Conditional expectation:

$$E_X[x] = E_Y[E_X[x|y]]$$

we have:

$$E_Y[E_X[x|y]] = \int_{-\infty}^{\infty} E_X[x|y] . f(y) . dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x . f_{X|Y}(x|y) . dx . f_Y(y) . dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x . f_{X|Y}(x|y) . f_Y(y) . dy . dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x . f_{X,Y}(x,y) . dy . dx$$

$$= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y) . dy \right) . dx$$

$$= \int_{-\infty}^{\infty} x . f_X(x) . dx$$

$$= E_X[X] \quad (q.e.d)$$

8. Inner product space and induced norm:

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad y = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

a) $\langle x, y \rangle := x^T y$

we have:

$$dist(x,y) = \|x - y\| = \sqrt{\langle x-y, x-y \rangle}$$

$$x - y = \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} \Rightarrow (x-y)^T = [2 \ 3 \ 3]$$

~~So dist(x,y) =~~

And: $\langle x-y, x-y \rangle = (x-y)^T (x-y)$

$$= [2 \ 3 \ 3] \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$$

$$= 2.2 + 3.3 + 3.3$$

$$= 22$$

So $dist(x,y) = \sqrt{22}$

b) $\langle x, y \rangle = x^T . A . y$

So $\langle x - y, x - y \rangle = (x-y)^T . A . (x-y)$

$$= \begin{bmatrix} 2 & 3 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 7 & 8 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$$

$= 7.2 + 8.3 + 3.3$

$= 47$

So $dist(x, y) = \sqrt{47}$

7. Chain rules:

$f(z) = e^{\left(-\frac{1}{2}.z\right)}$

$z = g(y) = y^T . S^{-1} . y$

$y = h(x) = x - \mu$

a) $\dfrac{\partial f}{\partial z} = \left(-\dfrac{1}{2} z\right)' e^{\left(-\frac{z}{2}\right)} = -\dfrac{1}{2} e^{\left(-\frac{z}{2}\right)}$

Dimension : $\mathbb{R}$

b) $\dfrac{\partial g}{\partial y} = y^T (S^{-1} + (S^{-1})^T)$

Dimension :

- We have $y \in \mathbb{R}^D \Rightarrow y^T \in \mathbb{R}^{1\times D}$

$\left.\begin{array}{l} S^{-1} \in \mathbb{R}^{D\times D} \\ (S^{-1})^T \in \mathbb{R}^{D\times D} \end{array}\right\}$ Dimension $= \mathbb{R}^{1\times D}$

c) $\frac{\partial h}{\partial x} = \frac{\partial(x-\mu)}{\partial x} = I - 0 = I$

with I is the $D \times D$ identity matrix

$\Rightarrow$ Dimension : $\mathbb{R}^{D \times D}$

d) ~~After~~ Using chain rule, we have

$$\frac{df}{dx} = \frac{df}{dz} \cdot \frac{dz}{dy} \cdot \frac{dy}{dx} = \frac{df}{dz} \cdot \frac{dg}{dy} \cdot \frac{dh}{dx}$$

$$= \frac{-1}{2} \cdot e^{\left(\frac{-z}{2}\right)} \cdot y^T \cdot (S^{-1} + (S^{-1})^T) \cdot I$$

Dimension : $\mathbb{R}^{1 \times D}$

9. Eigenvalues and eigenvectors :

$$A = \begin{bmatrix} 1 & 3 & 6 \\ 2 & 1 & 4 \\ 1 & 0 & 3 \end{bmatrix}$$

we have :

$$A \cdot \vec{v} = \lambda \cdot \vec{v}$$

$$\Rightarrow (A - \lambda I) \cdot \vec{v} = 0$$

$$\Rightarrow \det(A - \lambda I) = 0$$

$$A - \lambda I = \begin{bmatrix} 1-\lambda & 3 & 6 \\ 2 & 1-\lambda & 4 \\ 1 & 0 & 3-\lambda \end{bmatrix}$$

$$\Rightarrow \det(A - \lambda I) = 1 \cdot \det\begin{pmatrix} 3 & 6 \\ 1-\lambda & 4 \end{pmatrix} + 0 + (3-\lambda) \cdot \det\begin{pmatrix} 1-\lambda & 3 \\ 2 & 1-\lambda \end{pmatrix}$$

$$= 12 - 6(1-\lambda) + (3-\lambda)((1-\lambda)^2 - 6)$$

$$= -\lambda^3 + 5\lambda^2 + 5\lambda - 9$$

$$= (1-\lambda) \cdot (\lambda - 2 - \sqrt{13}) \cdot (\lambda - 2 + \sqrt{13}) = 0$$

$$\Rightarrow \left[\begin{array}{l} \lambda = 1 \\ \lambda = 2 + \sqrt{13} \\ \lambda = 2 - \sqrt{13} \end{array}\right.$$

So clearly $\lambda = 2 \pm \sqrt{13}$ cannot have integer vector as eigenvector

$\Rightarrow \lambda = 1$ is the only one can match a, b or c and only b satisfy

$$\begin{bmatrix} 1-1 & 3 & 6 \\ 2 & 1-1 & 4 \\ 1 & 0 & 3-1 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

So b is the correct answer

10. Eigenvalues and eigenvectors:

$$A = \begin{bmatrix} 1 & \lambda \\ 2 & 1 \end{bmatrix}$$

x is eigenvalue

$\Rightarrow \det(A - x.I) = 0$

$$\Rightarrow \det\left(\begin{bmatrix} 1-x & \lambda \\ 2 & 1-x \end{bmatrix}\right) = 0$$

$\Leftrightarrow (1-x)^2 - 2\lambda = 0$

$\Leftrightarrow (1-x)^2 = 2\lambda$

$\Leftrightarrow 1 - x = \pm\sqrt{2\lambda} \quad (\lambda \geq 0)$

$$\Leftrightarrow \left[\begin{array}{l} x = 1 + \sqrt{2\lambda} \\ x = 1 - \sqrt{2\lambda} \end{array}\right.$$

~~if $x = 1 + \sqrt{2\lambda}$~~

Eigenvector $= \begin{bmatrix} a \\ b \end{bmatrix}$

if $x = 1 + \sqrt{2\lambda}$

$(A - x.I).\vec{v} = 0$

$$\begin{bmatrix} -\sqrt{2\lambda} & \lambda \\ 2 & -\sqrt{2\lambda} \end{bmatrix} . \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{cases} -a.\sqrt{2\lambda} + b\lambda = 0 \Leftrightarrow a = \frac{\sqrt{\lambda}}{\sqrt{2}}.b \\ 2a - b\sqrt{2\lambda} = 0 \Leftrightarrow a = \frac{\sqrt{\lambda}}{\sqrt{2}}.b \end{cases}$$

$\Rightarrow$ eigenvector $= \begin{bmatrix} \frac{\sqrt{\lambda}}{\sqrt{2}}.t \\ t \end{bmatrix}$ with $\lambda, t \in \mathbb{R}$; $\lambda \geq 0$

if $x = 1 - \sqrt{2\lambda}$

$(A - x.I).\vec{v} = 0$

$$\begin{bmatrix} \sqrt{2\lambda} & \lambda \\ 2 & \sqrt{2\lambda} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{cases} a\sqrt{2\lambda} + b\lambda = 0 \Leftrightarrow a = b.\frac{-\sqrt{\lambda}}{\sqrt{2}} \\ 2a + b\sqrt{2\lambda} = 0 \Leftrightarrow a = b.\frac{-\sqrt{\lambda}}{\sqrt{2}} \end{cases}$$

$\Rightarrow$ eigenvector $= \begin{bmatrix} -\frac{\sqrt{\lambda}}{\sqrt{2}}.t \\ t \end{bmatrix}$ with $\lambda, t \in \mathbb{R}$; $\lambda \geq 0$

3. Conditional Independence:

$$\underbrace{(A \perp B \mid D)}_{(X)} \wedge \underbrace{(A \perp C \mid B, D)}_{(Y)} \Rightarrow \underbrace{(A \perp B, C \mid D)}_{(Z)}$$

(X): $P(A \perp B \mid D) = P(A \mid D).P(B \mid D)$

$$\Leftrightarrow \frac{P(A,B,D)}{P(D)} = \frac{P(A,D)}{P(D)} . \frac{P(B,D)}{P(D)}$$

$$\Leftrightarrow P(A,B,D) = \frac{P(A,D)}{P(D)}.P(B,D) \quad (1)$$

(Y): $P(A \perp C \mid B, D) = P(A \mid B, D) \cdot P(C \mid B, D)$

$$\Leftrightarrow \frac{P(A, C, B, D)}{P(B, D)} = \frac{P(A, B, D)}{P(B, D)} \cdot \frac{P(B, C, D)}{P(B, D)}$$

$$\Leftrightarrow P(A, B, C, D) = \frac{P(A, B, D)}{P(B, D)} \cdot P(B, C, D) \quad (2)$$

(Z): $P(A \perp B, C \mid D) = P(A \mid D) \cdot P(B, C \mid D)$

$$\Leftrightarrow \frac{P(A, B, C, D)}{P(D)} = \frac{P(A, D)}{P(D)} \cdot \frac{P(B, C, D)}{P(D)}$$

$$\Leftrightarrow P(A, B, C, D) = \frac{P(A, D)}{P(D)} \cdot P(B, C, D) \quad (3)$$

From (1)(2), we have:

$$P(A, B, C, D) = \frac{P(A, B, D)}{P(B, D)} \cdot P(B, C, D)$$

$$= \frac{P(A, D) \cdot P(B, D)}{P(D)} \cdot \frac{P(B, C, D)}{P(B, D)}$$

$$= \frac{P(A, D)}{P(D)} \cdot P(B, C, D) = (3)$$

So from (1) and (2) we have that (3) is true

so $(X) \wedge (Y) \Rightarrow (Z)$ (q.e.d)