Mohammadsadegh Abolhasani, u1416052
CS6190, Homework 0, Spring 2023

1. For the first one and according to the definition we have:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

Also, We know that the subset of two events will be non-negative. We can now move $p(A \cap B)$ to the other side of the equation and we will have:

$$\begin{aligned} &p(A \cup B) + p(A \cap B) = p(A) + p(B) \\ &p(A \cap B) \geq 0 \\ \rightarrow &p(A \cup B) \leq p(A) + p(B) \end{aligned}$$

Accordingly, the equity will hold if $p(A \cap B)$ equals zero, which will happen if the two events are independent (disjoint).
For the second one, we have to keep in mind that the probability of an event will be a real number between 0 and 1. Therefore, multiplying a non-negative number by the probability of an event will make it lesser than the original number:

$$\begin{aligned} &p(A \cap B) = p(A) * p(B|A) = p(B) * p(A|B) \\ &1 \geq p(A|B), p(B|A) \geq 0 \\ \rightarrow &p(A \cap B) \leq p(A), p(A \cap B) \leq p(B) \end{aligned}$$

And the equity of each of the expressions will be when A is a subset of B (for $p(A \cap B) = p(A)$) and when B is a subset of A (for $p(A \cap B) = p(B)$)

2. As hinted, I will use induction to prove this inequality. First, we will look at step one of this inequality:

$$p(A_1) \leq p(A_1) \tag{1}$$

Then, we will jump to the Nth step to see what the case will be:

$$p(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} p(A_i) \tag{2}$$

We will now try to implement the $n + 1th$ step. For that issue, we will need to review the union formula:

$$p(A_i \cup A_j) = p(A_i) + p(A_j) - p(A_i \cap A_j)$$

Then, we will add the $n+1st$ element to the inequality in step (2) for the next step. that shall be as follows:

$$p(\cup_{i=1}^{n+1} A_i) = p(\cup_{i=1}^{n} A_i) + P(A_{n+1}) - P(\cup_{i=1}^{n} + A_i \cap A_{n+1}) \tag{3}$$

From the previous question, we already know that the intersection between to events is non-negative. Therefore we have $P(\cup_{i=1}^{n} + A_i \cap A_{n+1}) \geq 0$ which will lead us to the following equation if we move this part to the left:

$$\begin{aligned} &p(\cup_{i=1}^{n+1} A_i) + P(\cup_{i=1}^{n} + A_i \cap A_{n+1}) = p(\cup_{i=1}^{n} A_i) + P(A_{n+1}) \\ \rightarrow &p(\cup_{i=1}^{n+1} A_i) \leq p(\cup_{i=1}^{n} A_i) + P(A_{n+1}) \qquad (4) \end{aligned}$$

Recalling (2) and (4) together we will have:

$$p(\cup_{i=1}^{n+1} A_i) \leq \sum_{i=1}^{n} p(A_i) + P(A_{n+1}) \rightarrow p(\cup_{i=1}^{n+1} A_i) \leq \sum_{i=1}^{n+1} p(A_i)$$

Which will prove the inequality in question. The equity will hold when all the events are independent with respect to each other so that every intersection will be zero.

3.

(a)

i. $P(x=0) = P(x=0, y=0) + P(x=0, y=1) = 3/10 + 1/10 = 4/10$
$P(x=1) = P(x=1, y=0) + P(x=1, y=1) = 2/10 + 4/10 = 6/10$
$P(y=0) = P(y=0, x=0) + P(y=0, x=1) = 3/10 + 2/10 = 5/10$
$P(y=1) = P(y=1, x=0) + P(y=1, x=1) = 1/10 + 4/10 = 5/10$

ii. According to the previous part and its calculations we have: $p(X|Y) = \frac{P(X \cap Y)}{P(Y)}$
$P(X=0|Y=0) = \frac{3/10}{5/10} = 3/5, P(X=1|Y=0) = \frac{2/10}{5/10} = 2/5$
$P(X=0|Y=1) = \frac{1/10}{5/10} = 1/5, P(X=1|Y=1) = \frac{4/10}{5/10} = 4/5$
$P(Y=0|X=0) = \frac{3/10}{4/10} = 3/4, P(Y=1|X=0) = \frac{1/10}{4/10} = 1/4$
$P(Y=0|X=1) = \frac{2/10}{6/10} = 1/3, P(Y=1|X=1) = \frac{4/10}{6/10} = 2/3$

iii. $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathbb{V}(X)$, $\mathbb{V}(Y)$
$\mathbb{E}(X) = 0 * P(x=0) + 1 * P(x=1) = 0 + 6/10 = 6/10$
$\mathbb{E}(Y) = 0 * P(y=0) + 1 * P(y=1) = 0 + 5/10 = 5/10$
$\mathbb{V}(X) = \Sigma(x - \mathbb{E}(x))^2 * P(x) \rightarrow \mathbb{V}(X) = ((0-0.6)^2 * 0.4) + ((1-0.6)^2 * 0.6) = 0.24$
$\mathbb{V}(Y) = ((0-0.5)^2 * 0.5) + ((1-0.5)^2 * 0.5) = 0.25$

iv. $\mathbb{E}(Y|X=0) = \Sigma y * P(Y|X=0) = 0 * 0.3 + 1 * 0.25 = 0.25$
$\mathbb{E}(Y|X=1) = 0 * 1/3 + 1 * 2/3 = 0.667$
$\mathbb{V}(Y|X=0) = \Sigma(y - \mathbb{E}(Y|X=0))^2 * P(Y|X=0) = (0-0.25)^2 * 3/4 + (1-0.25)^2 * 2/3 = 0.234$
$\mathbb{V}(Y|X=1) = \Sigma(y - \mathbb{E}(Y|X=1))^2 * P(Y|X=1) = (0-2/3)^2 * 3/4 + (1-2/3)^2 * 2/3 = 0.407$

v. $cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y), \mathbb{E}(XY) = \Sigma XY * P(XY), \mathbb{E}(XY) = 0*0*0.3 + 0*1*0.1 + 1*0*0.2 + 1*1*0.4 \rightarrow cov(X,Y) = 0.4 - 0.6*0.5 = 0.1$

(b) For them to be independent we should have $P(X_i, Y_J) = P_X(X_i) * P_Y(Y_j)$. But we see that doesn't hold true. (Like in $X=1, Y=1$ where we get $\frac{4}{10}$ instead of $\frac{5}{10} * \frac{6}{10}$)

(c) From the previous part, we realized that the two variables are dependent. Therefore, when X is not assigned a value, $\mathbb{E}(Y|X)$ and $\mathbb{V}(Y|X)$ will be functions of X and not constant.

4.

(a) $\mathbb{E}(Y) = \int_{-\infty}^{+\infty} yf(y) = \frac{1}{\sqrt{2\pi}} \int xe^{\frac{-x^2}{2}} e^{-x^2} = \frac{1}{\sqrt{2\pi}} \int xe^{\frac{-3x^2}{2}} = \frac{-1}{3\sqrt{2\pi}} e^{\frac{-3x^2}{2}} \Big|_{-\infty}^{+\infty} = \frac{1}{\sqrt{2\pi}} * \sqrt{\frac{2\pi}{3}} = 0.577$

(b) $\mathbb{V}(Y) = \int_{-\infty}^{+\infty} y^2 f(y) - (\mathbb{E}(Y))^2 = \frac{1}{\sqrt{2\pi}} \int e^{-2x^2} xe^{\frac{-x^2}{2}} e^{-x^2} - 0.577^2 = \frac{1}{\sqrt{2\pi}} \int xe^{\frac{-5x^2}{2}} - 0.577^2 = \frac{-1}{5\sqrt{2\pi}} e^{\frac{-5x^2}{2}} \Big|_{-\infty}^{+\infty} - 0.577^2 = 0.114$

(c) $\text{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \rightarrow \text{cov}(X,Y) = \mathbb{E}(xe^{-x^2}) - 0*577 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{\frac{-x^2}{2}} e^{-x^2} - 0 = \frac{-1}{3\sqrt{2\pi}} e^{\frac{-3x^2}{2}} \Big|_{-\infty}^{+\infty}$ (Even function) $-0 = 0$

5.

(a) $P_X(X) = \frac{1}{\sqrt{2*\pi}} * e^{\frac{-x^2}{2}}, Y = X^3 \rightarrow X = Y^{\frac{1}{3}} \rightarrow \frac{dx}{dy} = \frac{df^{-1}(Y)}{dY} = \frac{1}{3} Y^{\frac{-2}{3}} \rightarrow P_Y(Y) = P_X(f^{-1}(Y)) * \frac{df^{-1}(Y)}{dY} = \frac{1}{3\sqrt{2\pi}} y^{\frac{-2}{3}} e^{-\frac{y^{\frac{2}{3}}}{2}}$

(b)

6.

(a) $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(\Sigma_y yP(Y=y|X)) = \Sigma_X(\Sigma_y YP(Y=y|X=x))P(X=x) = \Sigma_X \Sigma_Y yP(Y=y, X=x) = \Sigma_Y y \Sigma_X P(Y=y, X=x) = \Sigma_Y yP(Y=y) = E(Y)$

(b) From definition of Variance, we have: $\mathbb{V}(Y|X) = \mathbb{E}(Y^2|X) - (\mathbb{E}(Y|X))^2 \rightarrow \mathbb{E}(\mathbb{V}(Y|X)) = \mathbb{E}(\mathbb{E}(Y^2|X)) - \mathbb{E}(\mathbb{E}(Y|X)^2)$, from the last part we know that $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y) \rightarrow \mathbb{E}(\mathbb{E}(Y^2|X)) = \mathbb{E}(Y^2) \rightarrow \mathbb{E}(\mathbb{V}(Y|X)) = \mathbb{E}(Y^2) - \mathbb{E}(\mathbb{E}(Y|X))^2$ (1) $\rightarrow \mathbb{V}(\mathbb{E}(Y|X)) = \mathbb{E}((\mathbb{E}(Y|X))^2) - (\mathbb{E}(Y))^2$ (2) $\rightarrow$ adding (1) and (2) we get: $\mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X)) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \mathbb{V}(Y)$

7.

(a) we first set u $u = 1 + e^{-\mathbf{a}^\top \mathbf{x}}$, then we see: $\frac{d}{d\mathbf{x}} f(x) = \frac{d}{du} u^{-1} \frac{du}{d\mathbf{x}} = -u^{-2} \frac{du}{d\mathbf{x}} \rightarrow \frac{d}{d\mathbf{x}} f(x) = -\frac{\frac{d}{d\mathbf{x}} u}{u^2} \rightarrow \frac{d}{d\mathbf{x}} f(x) = -\frac{\frac{d}{d\mathbf{x}}(1+e^{-\mathbf{a}^\top \mathbf{x}})}{(1+e^{-\mathbf{a}^\top \mathbf{x}})^2} = -\frac{\frac{d}{d\mathbf{x}} * 1 + \frac{d}{d\mathbf{x}} * e^{-\mathbf{a}^\top \mathbf{x}}}{(1+e^{-\mathbf{a}^\top \mathbf{x}})^2}$, $\frac{d}{d\mathbf{x}}(\mathbf{x}^\top \mathbf{a}) = \frac{d}{d\mathbf{x}}(\mathbf{a}^\top \mathbf{x}) = a \rightarrow \frac{d}{d\mathbf{x}} f(x) = \frac{a e^{-\mathbf{a}^\top \mathbf{x}}}{(1+e^{-\mathbf{a}^\top \mathbf{x}})^2}$

(b) We'll continue from the previous part by deriving the result w.r.t. x one more time. We will also us this rule: $(f * g)' = f' * g + g' * f$
$\frac{d^2 f(\mathbf{x})}{d\mathbf{x}^2} = \frac{d}{d\mathbf{x}}\left(\frac{a e^{-\mathbf{a}^\top \mathbf{x}}}{(1+e^{-\mathbf{a}^\top \mathbf{x}})^2}\right) = \frac{d}{d\mathbf{x}}((a e^{-\mathbf{a}^\top \mathbf{x}})((1+e^{-\mathbf{a}^\top \mathbf{x}})^{-2})) \rightarrow \frac{d^2 f(\mathbf{x})}{d\mathbf{x}^2} = (\frac{d}{d\mathbf{x}}(a e^{-\mathbf{a}^\top \mathbf{x}}))(1+e^{-\mathbf{a}^\top \mathbf{x}})^{-2} + (a e^{-\mathbf{a}^\top \mathbf{x}})) + (a e^{-\mathbf{a}^\top \mathbf{x}})(\frac{d}{d\mathbf{x}}(1 + e^{-\mathbf{a}^\top \mathbf{x}})^{-2}) = -\mathbf{a}\mathbf{a}^\top e^{-\mathbf{a}^\top \mathbf{x}}(1 + e^{-\mathbf{a}^\top \mathbf{x}})^{-2} + (a e^{-\mathbf{a}^\top \mathbf{x}}) * -2 * (0 - a e^{-\mathbf{a}^\top \mathbf{x}})(1 + e^{-\mathbf{a}^\top \mathbf{x}})^{-3} = \frac{-\mathbf{a}\mathbf{a}^\top e^{-\mathbf{a}^\top \mathbf{x}}}{(1+e^{-\mathbf{a}^\top \mathbf{x}})^2} + \frac{2\mathbf{a}e^{-2\mathbf{a}^\top \mathbf{x}}}{(1+e^{-\mathbf{a}^\top \mathbf{x}})^3}$

(c) To show that $-\log(f(\mathbf{x}))$ is convex, we have to derivate the $-\log(f(\mathbf{x}))$ function two times w.r.t. x:
$y = -log(f(\mathbf{x})) = -log(\frac{1}{1+exp(-\mathbf{a}^\top \mathbf{x})})$
$\frac{dy}{d\mathbf{x}} = \frac{d}{d\mathbf{x}}(-log(\frac{1}{1+exp(-\mathbf{a}^\top \mathbf{x})})) = \frac{d}{d\mathbf{x}}(log(1+exp(-\mathbf{a}^\top \mathbf{x}))) = \frac{1}{1+exp(-\mathbf{a}^\top \mathbf{x})} * \frac{d(-\mathbf{x}\mathbf{a}^\top)}{d\mathbf{x}} * exp(-\mathbf{a}^\top \mathbf{x})$, $\frac{d(-\mathbf{x}\mathbf{a}^\top)}{d\mathbf{x}} = \mathbf{a} \rightarrow \frac{dy}{d\mathbf{x}} = -a\frac{exp(-\mathbf{a}^\top \mathbf{x})}{1+exp(-\mathbf{a}^\top \mathbf{x})}$
The first round is complete. We should take the derivative one more time and see if the result is always positive:
$\frac{d^2 y}{d\mathbf{x}^2} = \frac{dy}{d\mathbf{x}}(-a\frac{exp(-\mathbf{a}^\top \mathbf{x})}{1+exp(-\mathbf{a}^\top \mathbf{x})}) = -a\frac{(\frac{d}{d\mathbf{x}} exp(-\mathbf{a}^\top \mathbf{x}))(1+exp(-\mathbf{a}^\top \mathbf{x})) - (\frac{d}{d\mathbf{x}}(1+exp(-\mathbf{a}^\top \mathbf{x})))(exp(-\mathbf{a}^\top \mathbf{x}))}{(1+exp(-\mathbf{a}^\top \mathbf{x}))^2}$
$\rightarrow \frac{d^2 y}{d\mathbf{x}^2} = -a\frac{(-a exp(-\mathbf{a}^\top \mathbf{x}))(1+exp(-\mathbf{a}^\top \mathbf{x})) - (-a(1+exp(-\mathbf{a}^\top \mathbf{x})))(exp(-\mathbf{a}^\top \mathbf{x}))}{(1+exp(-\mathbf{a}^\top \mathbf{x}))^2} = \frac{a^2 exp(-\mathbf{a}^\top \mathbf{x})}{(1+exp(-\mathbf{a}^\top \mathbf{x}))^2}$
Since the above sentence is always positive, therefore, $-log(f(x))$ is convex.

8. We define the convex conjugate as $g(\lambda) = max_x \lambda x - f(x)$, and to get the maximum, we should get $\lambda x - f(x)$ and then take its derivative and set it to be zero. This will give us the duality function.

(a) $f(x) = -\log(x), f^*(y) = max_{x \in domain(f)}(yx - f(x)), max: \frac{d}{dx}(yx - f(x)) = 0 \rightarrow \frac{d}{dx}(yx + log(x)) = y + \frac{1}{x} \rightarrow f^*(y) = -1 + log(-\frac{1}{y}) = -(1 + log(-y))$

(b) The same process as the last part:
$\frac{d}{dx}(y^\top \mathbf{x} - \mathbf{x}^\top A^{-1}\mathbf{x}) = d(\mathbf{y}^\top \mathbf{x}) - d(\mathbf{x}^\top \mathbf{A}^{-1}\mathbf{x}) = \mathbf{y}d\mathbf{x} - d(\mathbf{x}^\top \mathbf{A}^{-1}\mathbf{x}) = \mathbf{y}d\mathbf{x} - d\mathbf{x}^\top \mathbf{A}^{-1}\mathbf{x} - x^\top \mathbf{A}^{-1}d\mathbf{x}$
$d\mathbf{x}^\top \mathbf{A}^{-1}\mathbf{x}$ is a scalar and therefore it's its own transpose. We implement this together by knowing ${A^{-1}}^\top = A^{-1}$ to the equation and setting the result as $0 \rightarrow \mathbf{y}d\mathbf{x} - 2\mathbf{x}^\top \mathbf{A}^{-1}d\mathbf{x} = 0 \rightarrow \mathbf{x}^\top = \frac{1}{2}\mathbf{y}\mathbf{A} \rightarrow \mathbf{x} = \frac{1}{2}\mathbf{A}\mathbf{y}^\top \rightarrow f^*(y) = \mathbf{y}^\top \frac{1}{2}\mathbf{A}\mathbf{y}^\top - \frac{1}{2}\mathbf{y}\mathbf{A}\mathbf{A}^{-1}\frac{1}{2}\mathbf{A}\mathbf{y}^\top = \frac{1}{2}(\mathbf{A}\mathbf{y}^\top)(\mathbf{y}^\top - \frac{1}{2}\mathbf{y})$

9. [20 points] Derive the (partial) gradient of the following functions. Note that bold small letters represent vectors, bold captial letters matrices, and non-bold letters just scalars.

(a) We will use the definition of derivation to solve this one. The definition will see the behavior of $f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})$ for very small $\mathbf{h}$ values, where $\mathbf{h}$ is a vector. We will have:
$f(\mathbf{x}+\mathbf{h}) = (\mathbf{x}+\mathbf{h})^\top \mathbf{A}(\mathbf{x}+\mathbf{h}) = (\mathbf{x}^\top + \mathbf{h}^\top)\mathbf{A}(\mathbf{x}+\mathbf{h}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{x}^\top \mathbf{A}\mathbf{h} + \mathbf{h}^\top \mathbf{A}\mathbf{x} + \mathbf{h}^\top \mathbf{A}\mathbf{h} \rightarrow$
$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \mathbf{x}^\top \mathbf{A}\mathbf{h} + \mathbf{h}^\top \mathbf{A}\mathbf{x} + \mathbf{h}^\top \mathbf{A}\mathbf{h}$
$\rightarrow f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{h} + \mathbf{h}^\top \mathbf{A}\mathbf{x} + \mathbf{h}^\top \mathbf{A}\mathbf{h}$
By paying attention to the size of the variables and the output of the function, we can see that $\mathbf{h}^\top \mathbf{A}\mathbf{x}$ and $\mathbf{x}^\top \mathbf{A}\mathbf{h}$ are scalars. Therefore, they will be equal to their own transpose. And for $h^\top \mathbf{A}\mathbf{h}$, it will be limited to zero (since $\mathbf{h}$ is heading to zero) and can be neglected.
$\rightarrow f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{h} + \mathbf{x}^\top \mathbf{A}^\top \mathbf{h} + \mathbf{h}^\top \mathbf{A}\mathbf{h} \rightarrow \frac{\partial f}{\partial \mathbf{x}} = \mathbf{x}^\top \mathbf{A} + \mathbf{x}^\top \mathbf{A}^\top = \mathbf{x}^\top(\mathbf{A} + \mathbf{A}^\top)$

(b) We will change variables to make it easier. Also, due to size and output, we know that $A = (\mathbf{I} + \mathbf{x}\mathbf{x}^\top)^{-1}$ is a $n * n$ matrix.

$\partial f(\mathbf{x}) = \partial(\mathbf{A}^{-1}x) = \partial\mathbf{A}^{-1}\mathbf{x} + \mathbf{A}^{-1}\partial\mathbf{x}$
According to the rules: $\partial\mathbf{A}^{-1} = -\mathbf{A}^{-1}\partial\mathbf{A}\mathbf{A}^{-1}$
$\partial\mathbf{A} = \partial(\mathbf{I}+\mathbf{x}\mathbf{x}^\top) = \partial\mathbf{I} + \partial(\mathbf{x}\mathbf{x}^\top) = 0 + \partial\mathbf{x}\mathbf{x}^\top + \mathbf{x}\partial\mathbf{x}^\top$
Again, by paying attention to the size and the output, we can see that $\mathbf{x}^\top(\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1}$ is a scalar and therefore equals its transpose. Also, after finding out $\partial\mathbf{A}$, it's time to implement them into our equation:
$\partial f(x) = -\mathbf{x}^\top(\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1}\mathbf{x}(\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1}\partial\mathbf{x} - (\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1}\mathbf{x}\mathbf{x}^\top(\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1}\partial\mathbf{x} + (\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1}\partial\mathbf{x}$
$\rightarrow \frac{\partial f}{\partial\mathbf{x}} = -\mathbf{x}^\top(\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1}\mathbf{x}(\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1} - (\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1}\mathbf{x}\mathbf{x}^\top(\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1} + (\mathbf{I}+\mathbf{x}\mathbf{x}^\top)^{-1}$

(c) $f(\alpha) = \log|\mathbf{K}+\alpha\mathbf{I}|$, where $|\cdot|$ means the determinant. Derive $\frac{\partial f}{\partial\alpha}$.
We can change variables. We will take $\mathbf{A} = \mathbf{K}+\alpha\mathbf{I}$, and it's determinant to be $u = |\mathbf{K}+\alpha\mathbf{I}|$
$f(\alpha) = log(u)$, $\frac{\partial f}{\partial\alpha} = \frac{\partial log(u)}{\partial\alpha} * \frac{\partial u}{\partial\alpha} = \frac{1}{u}\frac{\partial u}{\partial\alpha}$
$\frac{\partial u}{\partial\alpha} = \frac{\partial|\mathbf{K}+\alpha\mathbf{I}|}{\partial\alpha} = \frac{\partial|A|}{\partial\alpha} = |A|\mathbf{Tr}(\mathbf{A}^{-1}\frac{\partial|A|}{\partial\alpha}) = \frac{\partial|A|}{\partial\alpha} = \frac{\partial(\mathbf{K}+\alpha\mathbf{I})}{\partial\alpha} = \frac{\partial\mathbf{K}}{\partial\alpha} + \frac{\partial\alpha\mathbf{I}}{\partial\alpha} = \mathbf{I}$
$\frac{\partial f}{\partial\alpha} = \frac{1}{|\mathbf{K}+\alpha\mathbf{I}|} * |\mathbf{K}+\alpha\mathbf{I}|\mathbf{Tr}(\mathbf{K}+\alpha\mathbf{I})^{-1} = \mathbf{Tr}(\mathbf{K}+\alpha\mathbf{I})^{-1}$

(d)

(e)

10. $p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = |2\pi\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left(-(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$, If we look at the equation, $x-\boldsymbol{\mu}$ leads us to change variables. suppose: $\mathbf{u} = \mathbf{x}-\boldsymbol{\mu} \rightarrow dx = d\mathbf{u} \rightarrow \frac{d}{d\mathbf{x}}p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{d}{d\mathbf{x}}|2\pi\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left(-(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) = \frac{d}{d\mathbf{u}}|2\pi\boldsymbol{\Sigma}|^{-\frac{1}{2}} * \frac{d}{d\mathbf{u}}\exp\left(-(\mathbf{u})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{u})\right)$
Then, according to last question's part (a) we know that $\frac{d}{d\mathbf{u}}\exp\left(\mathbf{u}^\top\boldsymbol{\Sigma}^{-1}\mathbf{u}\right) = \mathbf{u}^\top(\boldsymbol{\Sigma}^{-1}+(\boldsymbol{\Sigma}^{-1})^\top)$
$\rightarrow \frac{d}{d\mathbf{x}}p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = |2\pi\boldsymbol{\Sigma}|^{-\frac{1}{2}} * (\exp\left(\mathbf{u}^\top\boldsymbol{\Sigma}^{-1}\mathbf{u}\right))\mathbf{u}^\top(\boldsymbol{\Sigma}^{-1}+(\boldsymbol{\Sigma}^{-1})^\top)$
We know that this equation equals to zero if $\mathbf{u}^\top$ makes the sentence zero. Therefore, it should be a zero vector. Which according to our assumption back when we introduced the new variable means $\mathbf{u} = \mathbf{x}-\boldsymbol{\mu} = 0 \rightarrow \mathbf{x} = \boldsymbol{\mu}$

11. We can rewrite the left side of the formula this way to get better ideas. Now we have: $\int e^{-\frac{x^2}{2\sigma^2}}dx = \int e^{-(\frac{x}{\sqrt{2\sigma^2}})^2}dx$
We can now do a variable change to ease the integration. Take $u = \frac{x}{\sqrt{2\sigma^2}}$, it gives us: $du = \frac{1}{\sqrt{2\sigma^2}}dx \rightarrow dx = \sqrt{2\sigma^2}du \rightarrow \int e^{-(\frac{x}{\sqrt{2\sigma^2}})^2}dx = \int\sqrt{2\sigma^2}e^{-u^2}du$
We have to integrate the above equation to achieve the intended result. But first, by paying attention to the hint in the question, we will first square it. Also, by paying attention to the boundaries in the normal distribution, we know that it will have boundaries of $-\infty$ to ∞. We will take the integral part and square it and continue with our computation:
$(\int_{-\infty}^{\infty}e^{-u^2})^2 = \int_{-\infty}^{\infty}e^{-u^2}du * \int_{-\infty}^{\infty}e^{-v^2}dv = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-u^2-v^2}dudv = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(u^2+v^2)}dudv$
The u^2+v^2 part leads us to use polar coordinates, which will also show us where the π in the right hand of the equation in the question comes from. We shall now change variables and use polar coordinates:
$r^2 = u^2+v^2 \rightarrow dudv = rdrd\theta \rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(u^2+v^2)}dudv = \int_0^{2\pi}\int_0^{\infty}e^{-r^2}rdrd\theta \rightarrow \int_0^{2\pi}(-\frac{1}{2}e^{-r^2}\Big|_0^{\infty})d\theta = \frac{1}{2}\int_0^{2\pi}d\theta = \frac{1}{2}*2\pi = \pi$
Remember this was the result for the square of our initial integral. Therefore, we have to take the square root and place it in the initial equation. We will have:
$\int e^{-(\frac{x}{\sqrt{2\sigma^2}})^2}dx = \sqrt{2\sigma^2}\sqrt{\pi} = \sqrt{2\pi\sigma^2}$

12. As advised by the question, I will employ integral by parts, which is disclosed as $\int udv = uv - \int vdu$. In this case, we will select u in a way that makes the du and its calculation easier. Therefore, u^x is the option to go with for u. Subsequently, $du = xu^{x-1}, v = -e^{-u}, dv = e^{-u}$ and we will have the following:

$$\Gamma(x+1) = \int_0^{\infty}u^x e^{-u}\mathrm{d}u \rightarrow -u^x e^{-u}\Big|_0^{\infty} - \int -e^{-u}xu^{x-1}du \rightarrow x\int e^{-u}u^{x-1}du = x\Gamma(x)$$

Now we can solve the integral for $x = 1$, and we will have $\Gamma(1) = \int_0^\infty u^0 e^{-u} \mathrm{d}u$ which will be equal to $\int_0^\infty e^{-u} \mathrm{d}u$, The result will be: $-e^{-u}\Big| - 0^\infty = 1$

13. We know that $log(x)$ is concave since its second derivative is not positive. Therefore, I looked at the form of Jensen's inequality for concave functions and I came upon this in Wikipedia:
$\Phi = (\frac{\Sigma a_i x_i}{\Sigma a_i}) \geq \frac{\Sigma a_i \Phi(x_i)}{\Sigma a_i}$
Then, because we have nothing mentioned about the weights a_i, therefore we can conclude that they are equal and our inequality will transform to this new format:
$\Phi(\frac{\Sigma x_i}{n}) \geq \frac{\Sigma \Phi(x_i)}{n}$
Now we will implement our function, $log(x)$ in place of Φ, and we will have:
$log(\frac{\Sigma x_i}{n}) \geq \frac{\Sigma log(x_i)}{n}$
Taking the right hand of the equation will give us:
$\frac{\Sigma log(x_i)}{n} = \frac{1}{n}\Sigma_{i=1}^n log(x_i) = \Sigma_{i=1}^n \frac{1}{n} log(x_i) = \Sigma_{i=1}^n log(x_i^{\frac{1}{n}}) = log(\prod_{i=1}^n x_i^{\frac{1}{n}})$
Now coming back to our initial form of the inequality we will have:
$log(\frac{\Sigma_{i=1}^n x_i}{n}) \geq log(\prod_{i=1}^n x_i^{\frac{1}{n}}) \rightarrow \frac{1}{n}\Sigma_{i=1}^n x_i \geq (\prod_{i=1}^n x_i)^{\frac{1}{n}}$

14. In this question, we will deal with Jensen's inequality for convex functions since $-log(x)$ is convex.
$\mathbb{E}(f(x)) \geq f(\mathbb{E}(x)), f(x) = -log(x) \rightarrow \mathbb{E}(-log(\frac{p(x)}{q(x)})) \geq -log\mathbb{E}(\frac{p(x)}{q(x)}) = log\mathbb{E}(\frac{q(x)}{p(x)})$
Now if we want to compute the expectation based on $p(x)$ we will have:
$log\mathbb{E}(\frac{q(x)}{p(x)}) = log \int p(x)\frac{q(x)}{p(x)}dx = log \int p(x)dx$
Since $p(x)$ is a probability density function, its integration will be equal to 1 in its probability space. Therefore, we will have:
$log \int p(x)dx = log(1) = 0 \rightarrow \mathbb{E}(-log(\frac{p(x)}{q(x)})) \geq 0$

15.