problem 1

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

assume X and Y are indep.

1) want) $Var[X+Y] = Var[X] + Var[Y]$

$$
\begin{aligned}
Var[X+Y] &= E[(X+Y)^2] - (E[X+Y])^2 \\
&= E[X^2+Y^2+2XY] - (E[X]+E[Y])^2 \\
&= E[X^2] + E[Y^2] + 2E[XY] \\
&\quad - E[X]^2 - E[Y]^2 - 2E[X]E[Y] \\
&\quad (Y, X \text{ indep} \Rightarrow E[XY] = E[X]E[Y]) \\
&= (E[X^2] - E[X]^2) + (E[Y^2] - E[Y]^2) \\
&= Var[X] + Var[Y]
\end{aligned}
$$

2) want) $Var[XY] = E[X^2]E[Y^2] - E[X]^2 E[Y]^2$

$$
\begin{aligned}
var[XY] &= E[X^2Y^2] - (E[XY])^2 \\
&= E[X^2]E[Y^2] - (E[X]E[Y])^2 \\
&= E[X^2]E[Y^2] - E[X]^2 E[Y]^2
\end{aligned}
$$

3) want) if $E[X]=0$ $Var[XY] = Var[X]E[Y]^2$

$$
\begin{aligned}
var[XY] &= E[X^2]E[Y^2] - 0^2 \times E[Y]^2 \\
&= (E[X^2] - \underbrace{(E[X])^2}_{=0})E[Y^2] \\
&= Var[X]E[Y]^2
\end{aligned}
$$

problem 2.

1) Assume $C \in \mathbb{R}^n$ is a convex set and f and g both have C as their domain.

$f: C \to \mathbb{R}$ is convex means that

$$f(\gamma x + (1-\gamma)y) \le \gamma f(x) + (-\gamma) f(y)$$

$\forall x, y \in C$ and $\gamma \in [0,1]$.

$g: C \to \mathbb{R}^n$ is convex means that

$g(\gamma x + (1-\gamma) y) \le \gamma g(x) + (1-\gamma) g(y)$

$\forall x, y \in C$, $\gamma \in [0,1]$.

So we have

$\alpha f(\gamma x + (1-\gamma) y) + \beta g(\gamma x + (1-\gamma) y)$

$\le \alpha(\gamma f(x) + (1-\gamma) f(y)) + \beta(\gamma g(x) + (1-\gamma) g(y))$

$= \gamma(\alpha f(x) + \beta g(x)) + (1-\gamma)(\alpha f(y) - \beta g(y))$

$\forall x, y \in C$, $\gamma \in [0,1]$

$\Rightarrow \alpha f + \beta g$ is also convex func on C.

2) Assume $f: C \xrightarrow{\in \mathbb{R}^n} \mathbb{R}$ is convex.

<u>want</u>) f is both convex and concave

i.e. f and $-f$ are convex.

Let $\alpha \in [0,1]$ and $x, y \in C$.

Consider $f(\alpha x + (1-\alpha) y)$

$= \alpha f(x) + (1-\alpha) f(y) \le \alpha f(x) + (1-\alpha) f(y)$

(by property of linear map

i.e. if $f(x)$ is linear then

$f(c_1 u_1 + c_2 u_2) = c_1 f(u_1) + c_2 f(u_2)$

where $u_1, u_2 \in dom(f)$ $c_1, c_2 \in \mathbb{R}$ or $\mathbb{C}$)

$\Rightarrow f$ is convex on C

consider

$-f(\alpha x + (1-\alpha) y) = -(\alpha f(x) + (1-\alpha) f(y))$

$= \alpha(-f(x)) + (1-\alpha)(-f(y))$

$\le \alpha(-f(x)) + (1-\alpha)(-f(y))$

$\Rightarrow -f$ is convex on C

$\Rightarrow f$ is concave on C.

3) if $f(x)$ is a convex func on $\mathbb{R}^n$
we have
$f(\lambda x+(1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$
where $x, y \in \mathbb{R}^n$ $\lambda \in [0,1]$
consider
$g(\beta z + (1-\beta) h)$ $\beta \in [0,1]$ $z, h \in \mathbb{R}^m$
$= f(W[\beta z + (1-\beta) h] + b)$
$= f(\beta W z + (1-\beta) W h + \beta b + (1-\beta) b)$
$= f(\beta(Wz+b) + (1-\beta)(Wh+b))$
$\le \beta f(Wz+b) + (1-\beta) f(Wh+b)$
$= \beta g(z) + (1-\beta) g(h)$
$\Rightarrow g$ is convex on $\mathbb{R}^m$

problem 3

1) consider $g(t) = f(tx+(1-t)y)$
we have $g(1) = g(0) + g'(0) + \frac{1}{2} g''(\beta)$ ①
$g(1) = f(x)$
$g(0) = f(y)$
$g'(0) = \frac{d}{dt}(f(tx+(1-t)y)) \Big|_{t=0}$
$= f'(tx+(1-t)y)(x-y) \Big|_{t=0}$
$= f'(y)(x-y) = \nabla f(y)(x-y)$
$g''(\beta) = \frac{d^2}{d\beta^2}(f(\beta x + (1-\beta) y))$
$\cancel{= \frac{d}{d\beta}(f'(\beta x + (1-\beta) y)(x-y))}$
$= \frac{d}{d\beta}((x-y)^T f'(\beta x + (1-\beta) y))$
(To match the dimension
i.e. $\nabla f(y_\beta)$ is $n \times 1$ to get a result in
$\mathbb{R}$, we need $(x-y)^T$ $(1\times n)$).

$$\overset{\nearrow\, H(y_\beta)}{}$$

$$= \underset{1\times n}{(x-y)^T} \, \underset{n\times n}{f''(\beta x+(1-\beta)y)} \, \underset{n\times 1}{(x-y)} = 1\times 1 \checkmark$$

$$= (x-y)^T H(y_\beta)(x-y).$$

So ① $\Rightarrow$

$$f(x) = f(y) + \nabla f(y)(x-y) + \frac{1}{2}(x-y)^T H(y_\alpha)(x-y)$$

where $y_\alpha = \alpha x + (1-\alpha) y$, $\alpha \in [0,1]$,
(since β is arbitrary)

2) Assume f smooth in $\mathbb{R}^n$ and $v^T H v \geq 0$ $\forall v \in \mathbb{R}^n$

Then by previous result,

$$f(x) = f(y) + \nabla f(y)(x-y) + \underbrace{\frac{1}{2}(x-y)^T H(y_\beta)(x-y)}_{\geq 0 \text{ by assumption}},$$

$$\geq f(y) + \nabla f(y)(x-y)$$

$\Rightarrow$ $f(x)$ is convex on $\mathbb{R}^n$ by Lemma in the class.

3) $f(x) = g(h(x))$ $\quad f: \mathbb{R}^k \to \mathbb{R}$ ($1\times n$ matrix)

$$f'(x) = \underset{(1\times 1}{g'(h(x))} \, \underset{k\times 1}{h'(x)} = g'(h(x))(\nabla h(x)) \quad = k\times 1)$$

$$\underset{\nabla^2 f(x)}{=} f''(x) = \underset{(k\times k = 1\times 1}{g''(h(x))} \underset{k\times 1}{(\nabla h(x))} \underset{1\times k}{(h'(x))^T} + \underset{1\times 1}{g'(h(x))} \underset{k\times k)}{h''(x)}$$

$$= g''(h(x))(\nabla h(x))^T \nabla h(x)^T + g'(h(x)) \nabla^2 h(x).$$

4) $g(y) = \log y \qquad h(x) = \sum_{i=1}^{k} e^{x_i}$

$$f(x) = f(h(x)) = \log\left(\sum_{i=1}^{k} e^{x_i}\right)$$

$$g(h(x)) = \frac{1}{h(x)} \qquad g''(h(x)) = \frac{-1}{h^2(x)}$$

$$\nabla h(x) = \begin{pmatrix} e^{x_1} \\ e^{x_2} \\ \vdots \\ e^{x_k} \end{pmatrix} \qquad (\nabla h(x))^T = (e^{x_1}\ e^{x_2} \cdots e^{x_n})$$

$$\nabla^2 h(x) = \begin{pmatrix} e^{x_1} & & 0 \\ & e^{x_2} & \\ & & \ddots \\ 0 & & e^{x_k} \end{pmatrix}$$

(since e^{x_i} only involves the variable x_i

so $\frac{\partial h(x)}{\partial x_j \partial x_i} = 0 \quad \forall j \neq i$

combine

$$\Rightarrow \nabla^2 f(x) = -\frac{1}{h^2(x)} \begin{pmatrix} e^{x_1} \\ \vdots \\ e^{x_k} \end{pmatrix} (e^{x_1} \cdots e^{x_k}) + \frac{1}{h(x)} \begin{pmatrix} e^{x_1} & & 0 \\ & \ddots & \\ 0 & & e^{x_k} \end{pmatrix}$$

by previous result,

5) Let $v = \begin{pmatrix} v_1 \\ \vdots \\ v_k \end{pmatrix} \in \mathbb{R}^k$

for the $\nabla^2 f$ in 4), consider

① $v^T \begin{pmatrix} e^{x_1} \\ \vdots \\ e^{x_k} \end{pmatrix} (e^{x_1} \cdots e^{x_k}) v$

$$= \left((v_1, \cdots v_k) \begin{pmatrix} e^{x_1} \\ \vdots \\ e^{x_k} \end{pmatrix} \right) \left((e^{x_1} \cdots e^{x_k}) \begin{pmatrix} v_1 \\ \vdots \\ v_k \end{pmatrix} \right)$$

$$= \left(\sum_{i=1}^{k} v_i e^{x_i}\right)^2$$

② $v^T \begin{pmatrix} e^{x_1} & & 0 \\ & \ddots & \\ 0 & & e^{x_K} \end{pmatrix} v$

$$= v^T \left(\begin{pmatrix} e^{x_1} & & 0 \\ & \ddots & \\ 0 & & e^{x_K} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_K \end{pmatrix} \right)$$

$$= (v_1 \cdots v_K) \begin{pmatrix} e^{x_1} v_1 \\ \vdots \\ e^{x_K} v_K \end{pmatrix}$$

$$= \sum_{i=1}^{K} v_i^2 e^{x_i}$$

combine ①, ② and result in 4) we get

$$v^T H(x) v = -\frac{\left(\sum_{i=1}^{K} x_i e^{x_i}\right)^2}{h^2(x)} + \frac{\sum_{i=1}^{K} v_i^2 e^{x_i}}{h(x)}$$

6) use cauchy-schwatz inequality

i.e. $(\sum ab)^2 \le \sum a^2 \sum b^2$

for $\left(\sum_{i=1}^{K} v_i e^{x_i}\right)^2$ Let $a = e^{\frac{x_i}{2}}$ $b = v_i e^{\frac{x_i}{2}}$

then $\left(\sum_{i=1}^{K} v_i e^{x_i}\right)^2 \le \sum_{i=1}^{K} e^{x_i} \sum_{i=1}^{K} v_i^2 e^{x_i}$

$$\le h(x) \sum_{i=1}^{K} v_i^2 e^{x_i} \quad ①$$

Also, since $h(x) = \sum_{i=1}^{K} e^{x_i} > 0$ $\forall x \in \mathbb{R}^K$

so we have $1/h(x) > 0$.

Then

$$v^T H(x) v = \frac{-\left(\sum v_i e^{x_i}\right)^2}{h^2(x)} + \frac{\sum v_i^2 e^{x_i}}{h(x)}$$

$$= \frac{-\left(\sum v_i e^{x_i}\right)^2}{h^2(x)} + \frac{h(x) \sum v_i^2 e^{x_i}}{h^2(x)} \ge 0,$$

by ①

7) since $v^T \nabla^2 f(x) v \geq 0 \quad \forall v \in \mathbb{R}^k$
by 2), we get $f(x)$ convex on $\mathbb{R}^k$.

8) since $f(x) = \log\left(\sum_{i=1}^{k} e^{x_i}\right)$ is convex, on $\mathbb{R}^k$
then define
$$g(X) = f(WX + b) \text{ is convex on } \mathbb{R}^d$$
i.e. $W = \begin{pmatrix} w_1 \\ \vdots \\ w_k \end{pmatrix}$ where $w_i \in \mathbb{R}^d$ i.e. $\dim W = k \times d$

$b = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix} \in \mathbb{R}^k$ ($\dim b = k \times 1$)

$\dim X = d \times 1$ $\qquad g(X) : \mathbb{R}^d \to \mathbb{R}^k$.

so define $f(\theta) = g(x) = \log\left(\sum_{i=1}^{k} e^{W_i X + b_i}\right)$
where $\theta = \begin{pmatrix} w_1 \\ b_1 \\ \vdots \\ w_k \\ b_k \end{pmatrix}$ is convex on $\mathbb{R}^n$
where $n = k(d+1)$