$$-\tfrac{1}{2}(x^2-2x+1) = \tfrac{1}{2}(x-1)^2$$

$$-\tfrac{1}{2}x^2+2x = \tfrac{-1}{2}(x^2-4x+4) = -\tfrac{1}{2}(x-2)^2$$

c) $X \sim N(0,1)$. $Y = e^X$

$$\mathbb{E}[Y] = \mathbb{E}[e^x] = \int_{-\infty}^{\infty} e^x \left(\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}\right)dx = \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}x^2+x}}{\sqrt{2\pi}}dx$$

$$= \frac{e^{\frac{1}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2+x-\frac{1}{2}}dx = \frac{e^{1/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-1)^2}dx$$

$$= \frac{e^{1/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2}du \qquad u = x-1 \quad du = dx$$

$$= \frac{e^{1/2}}{\cancel{\sqrt{2\pi}}}\cdot \cancel{\sqrt{2\pi}} \quad \text{because of } \mathbb{E}[N]$$

$$= e^{1/2}$$

$$\mathbb{E}[Y]^2 = e^1$$

We need $\mathbb{E}[Y^2] = \int_{-\infty}^{\infty} e^{2x}\left(\frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}\right)dx$

$$= \frac{e^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2+2x-2}dx = \frac{e^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-2)^2}dx$$

So, $\mathrm{Var}[Y] = \underbrace{\mathbb{E}[Y^2]}_{=e^2} - \mathbb{E}(Y)^2 = e^2 - e^1$

$\mathbb{E}[Y] = e^{1/2}$

d) We know $\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}[Y]^2$.
So we need to check this against $\mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}\,\mathbb{E}(Y|X)$

$$\mathbb{E}[\mathrm{Var}[Y|X]] + \mathrm{Var}\,\mathbb{E}[Y|X] = \mathbb{E}[\mathbb{E}[Y^2|X] - \mathbb{E}[Y|X]^2] + (\mathbb{E}[(\mathbb{E}(Y|X))^2] - \mathbb{E}[\mathbb{E}(Y|X)]^2)$$

$$= \mathbb{E}[\mathbb{E}[Y^2|X]] - \mathbb{E}[\mathbb{E}[Y|X]^2] + \mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[\mathbb{E}[Y|X]]^2$$

$$= \mathbb{E}[Y^2] - \mathbb{E}(Y)^2 = \mathrm{Var}(Y). \checkmark$$

2. $\hat{y} = X\hat{\beta}$ where $\hat{y} = Hy$ and $H = X(X^TX)^{-1}X^T$

a) $\hat{y} = Hy = X\hat{\beta}$ are the least squares estimates.

$\hat{\beta}$ minimizes $\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta^T x_i)^2 \Rightarrow \min_{\beta} \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta^T x_i)^2$

$$\Rightarrow \min_{\beta} \frac{1}{n}(Y - X\beta)^T(Y - X\beta) = \frac{1}{n}\min_{\beta} y^Ty - \beta^TX^Ty - y^TX\beta + \beta^TX^TX\beta$$

$$= \frac{1}{n}\min_{\beta} y^Ty - 2\beta^TX^Ty + \beta^TX^TX\beta$$

$$= -2X^Ty + X^TX\hat{\beta} + \hat{\beta}^TX^TX = 0$$

$$= -2X^Ty + 2X^TX\hat{\beta} = 0$$

Note: $\beta^TX^Ty = y^TX\beta$, $X^TX\beta = \beta^TX^TX$

$$= -X^Ty + X^TX\hat{\beta} = 0$$

$$\Rightarrow X^Ty = X^TX\hat{\beta}$$

X^TX is non-singular $\Rightarrow \hat{\beta} = (X^TX)^{-1}X^Ty$

So, $X\hat{\beta} = X(X^TX)^{-1}X^Ty = Hy$

So, we have shown that these are the least squares estimates.

b) $HX = X \Rightarrow (X(X^TX)^{-1}X^T)X = X(X^TX)^{-1}(X^TX) = X(I) = X.$

c) We wish to show H is symmetric.

$H = H^T \Rightarrow H^T = (X(X^TX)^{-1}X^T)^T = (X^T)^T((X^TX)^{-1})^TX^T$

$$= X((X^TX)^{-1})^TX^T$$

Note: $(X^TX)^{-1}(X^TX) = I$

$((X^TX)^{-1}(X^TX))^T = I^T$

$(X^TX)^T((X^TX)^{-1})^T = I$

$X^TX((X^TX)^{-1})^T = I \Leftrightarrow ((X^TX)^{-1})^T = (X^TX)^{-1}$

So, $\Rightarrow X(X^TX)^{-1}X^T = H$

So, $H^T = H$

d) $H^2 = H$

$H^2 = (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T)$

$= X(X^TX)^{-1}\cancel{(X^TX)}\cancel{(X^TX)^{-1}}X^T$

$= X(X^TX)^{-1}X^T = H$

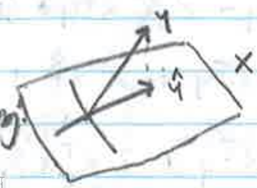
$1 \times d$

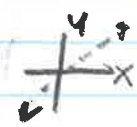
e) $\hat{y} = Hy$ is the projection of y onto $\mathcal{J} = \alpha_1 X_1 + \dots + \alpha_d X_d$

where x_i are columns of X.

$\hat{y} = \alpha_1 x_1 + \dots + \alpha_d x_d = X\alpha$

$\hat{y} = X\alpha = Hy$





$X^T(y - \hat{y}) = 0$, definition of proj

$\Rightarrow X^T(y - X\alpha) = 0$

$X^Ty - X^TX\alpha = 0$

$X^Ty = X^TX\alpha \Rightarrow (X^TX)^{-1}X^Ty = \alpha$

$\Rightarrow X(X^TX)^{-1}X^Ty = X\alpha = \hat{y}$

$\Rightarrow y = Hy$

So $\hat{y}$ is a projection of y onto $\mathcal{J}$.

f) $\mathrm{Rank}(X) = tr(H) = d$

$tr(H) = tr(X(X^TX)^{-1}X^T)$

$= tr(X^TX(X^TX)^{-1})$ by $tr(AB) = tr(BA)$

$tr(I) = d$

where I is a $d \times d$ matrix

We need $\mathrm{Rank}(X) = d$.

$\mathrm{Rank}(X) \le d$ since X is an $n \times d$ matrix and $d < n$.

$\det(X^TX) \neq 0$ since X^TX is nonsingular.

$\mathrm{Rank}(X^TX) = d$ since X^TX $d \times d$ matrix.

Consider, $X^T : X \to V$.

Since $\mathrm{rank}(X^T) = d$ in this case

we have $\mathrm{rank}(X) = d$.

So, $\mathrm{Rank}(X) = Tr(H) = d$.

3 a) $X = U\Sigma V^T$ $\quad X^T = V\Sigma^T U^T$ $\quad u_i = Ue_i$ $\quad v_i = Ve_i$ $\quad e_i$ are stand. basis vectors.

Let u_i be columns of U & v_i be columns of V.

If u_i are eigenvectors of XX^T then

$(XX^T)u_i = \lambda u_i \quad \lambda \in \mathbb{R}$

$\Rightarrow (U\Sigma V^T)(V\Sigma^T U^T)u_i$ by transpose

$\Rightarrow (U\Sigma\Sigma^T U^T)Ue_i$ by orthogonality

$= U\Sigma\Sigma^T e_i$

$= \sigma_i^2 Ue_i = \sigma_i^2 u_i = XX^T u_i$ so $\lambda = \sigma_i^2$ if $i \leq r$.

If v_i are eigenvectors of X^TX then

$(X^TX)v_i = \lambda v_i \quad \lambda \in \mathbb{R}$

$(X^TX)v_i = (V\Sigma^T U^T)(U\Sigma V^T)v_i$

$= V\Sigma^T I \Sigma V^T \cdot Ve_i$

$= (V\Sigma^T\Sigma)e_i$

$= \sigma_i^2 Ve_i = \sigma_i^2 v_i = X^TXv_i$

So, $\lambda = \sigma_i^2$ if $i \leq r$

b) We wish to show that $Xv_i = \sigma_i u_i$

$Xv_i = U\Sigma V^T V e_i = U\Sigma e_i = U\sigma_i e_i = \sigma_i Ue_i$

$= \sigma_i u_i$ by orthogonality.

Similarly $X^T u_i = (V\Sigma^T U^T)Ue_i = V\Sigma^T e_i = V\sigma_i e_i$

$= \sigma_i Ve_i = \sigma_i v_i$.

c) $\|X\|_F = \sqrt{\sum_{i,j} x_{ij}^2}$

$x_{ij} \quad x_{ji}$

$= \sqrt{\sum_i \sum_j x_{ij}^2}$ $\quad$ using $tr(AB) = \sum_{i,j=1}^{n} a_{ij} b_{ji}$

$= \sqrt{tr(XX^T)}$

$= \sqrt{tr(\cancel{U}\Sigma \cancel{V^T} \cancel{V} \Sigma^T \cancel{U^T})}$

$= \sqrt{tr(\Sigma\Sigma^T)}$

$= \sqrt{\sum_{i=1} \sigma_i^2}$

d) $|\det(X)| = |\det(U) \cdot \det(\Sigma) \cdot \det(V^T)|$

$|(\pm 1) \det(\Sigma) (\pm 1)|$ by orthogonality

$= |\det(\Sigma)|$

$= |\prod_i \sigma_i|$ since Σ is diagonal

e) $H = X(V\Sigma^T\Sigma V^T)^{-1} X^T$ since $X^TX = V\Sigma^T U^T U \Sigma V^T$

$= U\Sigma \cancel{V}^T(\cancel{V}(\Sigma^T\Sigma)^{-1}\cancel{V^T}) \cancel{V}\Sigma^T U^T$ by orthogonality

$= U\Sigma(\Sigma^T\Sigma)^{-1}\Sigma^T U^T$

f) $\hat{y} = X(X^TX)^{-1}X^T$

$= U\Sigma^{(k)}V^T (V\Sigma^{(k)T}U^TU\Sigma^{(k)}V^T)^{-1} V\Sigma^{(k)T}U^T$

$= U\Sigma^{(k)}(\Sigma^{(k)T}\Sigma^{(k)})^{-1}\Sigma^{(k)T}U^T$