1. $\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$

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

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

2. $\begin{bmatrix} 1 & 2 & 3 \\ -1 & 4 & 2 \\ 0 & 6 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 6 & 5 \\ 0 & 6 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 6 & 5 \\ 0 & 0 & 0 \end{bmatrix}$

Col space : span $\{[1,-1,0]^T, [2,4,6]^T\}$

Null Space : span $\{[-4/3, -5/6, 1]\}$

b) No : $AB \neq BA$ for 2 matrices A, B all the time
↳ matrix multiplication is not commutative

c) No : counter example :

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

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

To minimize $\|Ax - b\|$

we have : $Ax = b$

$A^T A x = A^T b$

$x = (A^T A)^{-1} A^T b$

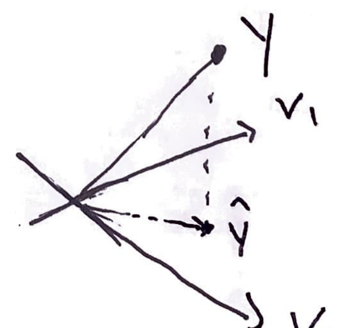
$*$ $A^T A$ is invertible

$b = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$

recall $A^{-1} = \dfrac{1}{|A|} \begin{bmatrix} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & \cdot & \cdot \\ \begin{vmatrix} a_{23} & a_{21} \\ a_{33} & a_{31} \end{vmatrix} & \cdot & \cdot \\ \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} & \cdot & \cdot \end{bmatrix}$

$\vec{x} = \begin{bmatrix} 6 \\ -1 \\ 3/2 \end{bmatrix}$

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



c) distance = actual − proj

$= \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix} - \begin{bmatrix} 4 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

4) $L(w) = \|Xw - y\|_2^2 + \lambda \|w\|_2^2$

$$= (Xw - y)^T (Xw - y) + \lambda w^T w$$

$$= (w^T X^T X w - w^T X^T y - y^T X w + y^T y) + \lambda w^T w$$

$$= (w^T X^T X w - 2 w^T X^T y + y^T y) + \lambda w^T w$$

$$\frac{\partial L}{\partial w} = 2X^T X w - 2X^T y + 2\lambda w = 0$$

↳ recall derivatives: $x^T B x \rightarrow 2Bx$

* $x^T x \rightarrow 2x$
* $x^T b \rightarrow b$

$$w = (\cancel{2} X^T X + \cancel{2}\lambda I)^{-1} \cancel{2} X^T y$$

$$\boxed{w = (X^T X + \lambda I)^{-1} X^T y}$$

The hyperparameter λ controls regularization: a higher λ means higher penalty & less extreme / smaller model weights (this helps in cases of overfitting). The smaller λ is, the closer we get to normal least-squares regression (aka $\lambda = 0$).