(07)

$\|AX - Y\|_2^2$ → Minimize.

⇒ Taking transpose. (norm remains same on transpose)

$\|X^T A^T - Y^T\|_2^2$ [A → a m×n matrix]

Let's call A^T as W

$\|X^T W - Y^T\|_2^2$.

$$\|X^T W - Y^T\|_2^2 = \sum_{i=1}^{n} \|X^T w_i - y_i\|_2^2$$

Each w_i appears only once so optimal w_i minimises

$\|X^T w_i - y_i\|_2^2$.

$$\hat{w}_i = (XX^T)^{-1} X y_i$$

Therefore

$$\hat{A} = \hat{W}^T$$

$$= \begin{bmatrix} (XX^T)^{-1} X y_1 \\ \vdots \\ (XX^T)^{-1} X y_n \end{bmatrix}$$

$$= ((XX^T)^{-1} X Y^T)^T$$

$$= YX^T (XX^T)^{-1}$$

$[X^T (XX^T)^{-1} = (X^T X)^{-1} X^T$ if rows independent]

$$\boxed{\hat{A} = YX^+}$$