On veut trouver le modèle linéaire qui relie observations X à un résultat observé Y

On prend en compte l'erreur / loss / énergie du système E

$$Y = X * A + E$$

↑ modèle linéaire

L'erreur / loss, c'est E d'où $\|E\|^2$ à minimiser

soit $\|Y - X*A\|^2 = {}^t(Y - XA)(Y - XA)$

(produit scalaire du vecteur $E = Y - XA$)

C'est ce qu'on appelle l'estimation par optimisation de l'erreur au moindre carré (MSE: Mean Squared Error)

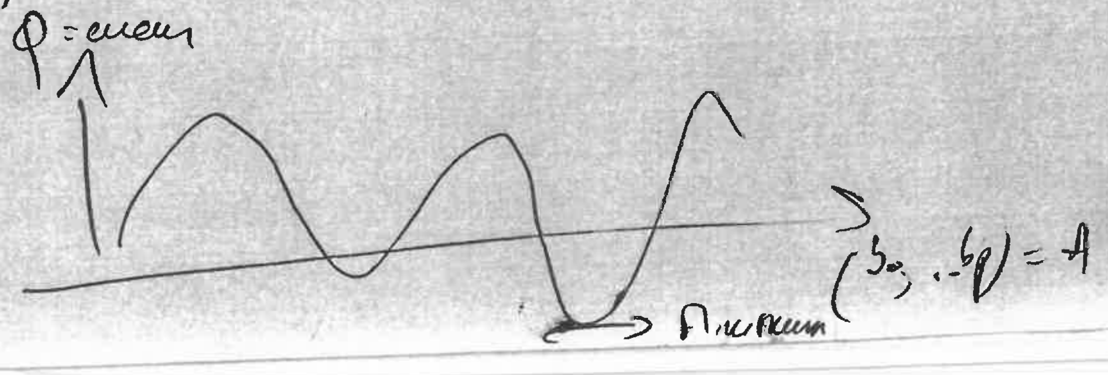
On peut montrer par Algèbre linéaire directement que le modèle appris est $A = ({}^tXX)^{-1}\,{}^tXY$

On peut aussi le faire par optimisation: trouver le minimum d'une fonction de cost / erreur / loss. (Principe de Machine Learning).

L'erreur s'écrit comme une fonction des paramètres du modèle $A = \begin{pmatrix} b_0 \\ \vdots \\ b_p \end{pmatrix}$:

$$\phi(b_0, b_1, \dots, b_p) = {}^t(Y - XA)(Y - XA)$$

On cherche à faire varier tous les paramètres $b_0, \dots$ à b_p pour minimiser cette erreur



Cela implique que

$$\forall j, \quad \frac{\partial \varphi}{\partial b_j} = 0$$

Exercice 2 :

$$\varphi(b_1,\dots,b_p) = {}^tYY - {}^tYXA - {}^tA\,{}^tXY + {}^tA\,{}^tX\,XA$$

donc $$\frac{\partial\varphi}{\partial b_j} = \cancel{\frac{\partial\,{}^tYY}{\partial b_j}} - {}^tYX\frac{\partial A}{\partial b_j} - \frac{\partial}{\partial b_j}{}^tA\,{}^tXY + \frac{\partial}{\partial b_j}({}^tA\,{}^tXXA)$$

Si U, V vecteurs de dimension identique, ${}^tUV$ = scalaire

donc ^{T}scalaire = scalaire (produit scalaire) donc

$${}^tUV = {}^tVU$$

donc $$\frac{\partial}{\partial b_j}{}^tA\,{}^tXY = {}^tYX\frac{\partial A}{\partial b_j}$$

et $$\frac{\partial}{\partial b_j}({}^tA\,{}^tX\,XA) = \frac{\partial}{\partial b_j}({}^t(XA)\,XA)$$

$$= 2\,{}^t\left(\frac{\partial XA}{\partial b_j}\right)XA = 2\,{}^t(XA)\frac{\partial XA}{\partial b_j}$$

$$= 2({}^t(XA)X\frac{\partial A}{\partial b_j}$$ (car $\frac{\partial X}{\partial b_j} =$

On met tout ensemble

$$-{}^tYX\frac{\partial A}{\partial b_j} - {}^tYX\frac{\partial A}{\partial b_j} + 2\,{}^tA\,{}^tXX\frac{\partial A}{\partial b_j} = 0 \quad \forall j$$

soit $$\frac{\partial A}{\partial b_j}(-{}^tYX - {}^tYX + 2\,{}^tA\,{}^tXX) = 0 \quad \forall j$$

soit $$(-2\,{}^tYX + 2\,{}^tA\,{}^tXX)^t = 0$$

soit $$-2X^tY + 2X^tXA = 0 \quad \forall j$$

$$\Rightarrow X^tY = X^tXA \Rightarrow A = (X^tX)^{-1}X^tY$$ CQFD.