Data Science Lab

Feb 6th — Lecture No6.

Today
- Variance, Covariance, Correlations
- Some fundamentals on Regression.

Random Variables

You have PMF or PDF

$$\mu = \mathbb{E}[X] = \sum_x x \cdot \underbrace{\mathbb{P}(X=x)}_{\text{PMF.}}$$

Samples

You have samples

$X_1, X_2, \ldots X_n$ iid Realizations of a R.V. X

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{sample mean.}$$

X_i are RVs.

$\bar{X}$ is a R.V.

$$E[\bar{X}] = \frac{1}{n} \cdot n \cdot \mu = \mu.$$

Since the expectation of the estimator $\bar{X}$ is indeed what we want to estimate, $\bar{X}$ is called an unbiased estimator.

another estimator for μ is $\bar{X}_2 = X_1$

$$E[\bar{X}_2] = \mathbb{E}(X_1) = \mu.$$

$$\bar{X}_3 = \frac{1}{2}(X_1 + X_2)$$

$$E(\bar{X}_3) = \frac{1}{2} \cdot 2\mu = \mu.$$

These all are unbiased estimators of μ. but they are not great.

$\mathrm{Var}(\bar{X})$.

Var of a R.V X.

$$= \mathbb{E}[\underbrace{(X - \mathbb{E}(X))^2}_{\text{squared deviations around the mean.}}] = \sigma^2.$$

$$\mathrm{Var}(\bar{X}) = \frac{1}{n^2} \cdot n \cdot \sigma_x^2 = \frac{1}{n} \cdot \sigma_x^2.$$

Recall
- $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ if X, Y indep.

$$\boxed{\text{If you double you samples } n \to 2n \text{ your error std deviation } \sigma \text{ goes down by } \frac{1}{\sqrt{2}} = \frac{1}{1.414..}}$$

Covariance of two R.Vs X, Y

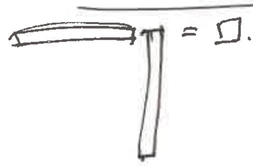
$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[\left(X - \mathbb{E}(X)\right)\cdot\left(Y - \mathbb{E}(Y)\right)\right].$$

Sanity check $\mathrm{Cov}(X,X) = \mathbb{E}\left[\left(X - \mathbb{E}(X)\right)\left(X - \mathbb{E}(X)\right)\right]$

$$= \mathbb{E}\left(\left(X - \mathbb{E}(X)\right)^2\right) = \mathrm{Var}(X).$$

<u>Covariance Matrix</u> of a Random Vector X.

$$\mathbb{E}\left[\left(X - \mathbb{E}(X)\right)\left(X - \mathbb{E}(X)\right)^T\right].$$

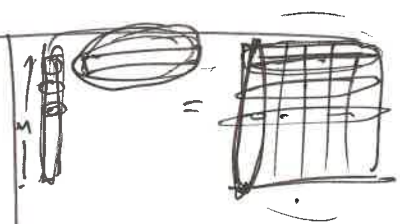


$X^T X$

scalar

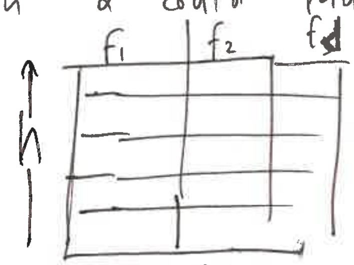
dot prod

inner prod.



$X X^T$.

if we are given samples in a data matrix



$\mathbb{D}$

<u>Sample covariance matrix</u>

Def: $\dfrac{1}{n-1} D_{cent}^T \cdot D_{cent}$

Correlation coefficient

$$\boxed{\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}}.$$

Ex $D =$

Weight	Height
80	160
85	170
75	180

estimate

Var W, Var H

St. dev. σ_W, σ_H

Covariance $Cov(W, H)$

Correlation coefficient

$\rho_{w,h}$.

Sample variance X

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2.$$

$\bar{w} = 80$, $s_w^2 = \frac{1}{2} \cdot (0^2 + 5^2 + 5^2) = 25$.

Sample $Cov(w, h) = \frac{1}{n-1} \sum (w_i - \bar{w}) \cdot (h_i - \bar{h}))$.

$$S_{wh}^2 = -25.$$

$S_h^2 = 100$.

$$\underbrace{\text{Sample Cov Matrix}} = \begin{bmatrix} 25 & -25 \\ -25 & 100 \end{bmatrix}.$$

Using Matrix calculations:

Center D matrix: $D_c = \begin{bmatrix} 0 & -10 \\ +5 & 0 \\ -5 & +10 \end{bmatrix}$.

Compute $\frac{1}{n-1} D_c^T \cdot D_c = \frac{1}{2} \begin{bmatrix} 0 & +5 & -5 \\ -10 & 0 & +10 \end{bmatrix} \cdot \begin{bmatrix} 0 & -10 \\ +5 & 0 \\ -5 & +10 \end{bmatrix}$

$$= \begin{bmatrix} 25 & -25 \\ -25 & 100 \end{bmatrix}.$$

numpy . np.cov(D^T)

Pandas . pd.df.cov(D).

4

$$\rho_{xy} = \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y}.$$ correlation coefficient.

For height, weight: $\sigma_w = 5$, $\sigma_h = 10$.

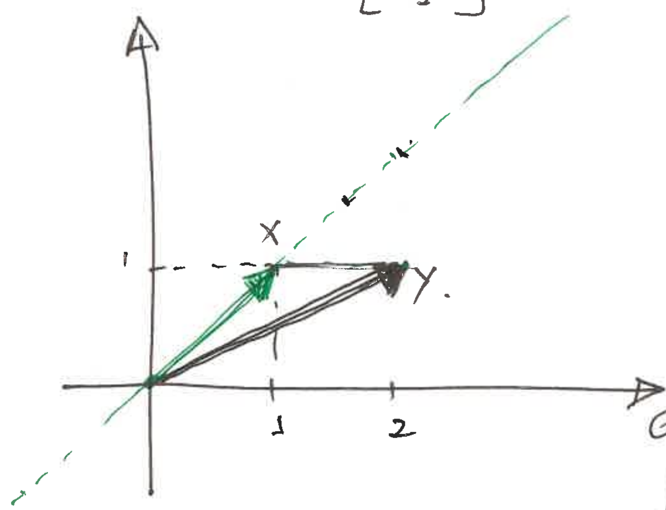
$$\rho_{wh} = \frac{-25}{5 \cdot 10} = -0.5.$$

ρ is always between $[-1, 1]$.

Lets do a super simple Linear Algebra question:

$\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\vec{y} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

I want to project Y on X.



I can create any vector $\beta \cdot \vec{x}$

cost $dist(\beta \cdot \vec{x}, \vec{y})$.

$$\min_{\beta} \|\beta \cdot \vec{x} - \vec{y}\|_2$$

Fundamental Fact: this is minimized when the error is orthogonal to the things you can express

$$\cancel{(\vec{y} - \beta \cdot \vec{x})^T \cdot (\beta \cdot x) = 0}.$$

$$\Rightarrow (\beta^* x)^T (y - \beta^* x) = 0 \Rightarrow$$

$$\cancel{\beta} \cdot x^T \cdot y = x^T \cdot x \cdot \beta^{*\cancel{2}} \Rightarrow.$$

$$x^T \cdot x \cdot \beta^* = x^T \cdot y. \Rightarrow \boxed{\beta^* = \frac{x^T \cdot y}{x^T \cdot x}.}$$

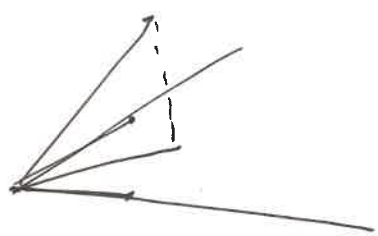
$$= \frac{x^T \cdot y}{\|x\|_2^2}$$

$$\beta^* = \frac{X^T \cdot y}{X^T \cdot X}.$$

The Remarkable thing is that this Formula is essentially correct even if you are projecting a vector y on a bunch of vectors x_1, x_2.

$$y = \begin{bmatrix}1\\2\\3\end{bmatrix} \qquad x_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \quad x_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}.$$

Project y on span of x_1, x_2.



$$\min_{\beta_1, \beta_2} \| \vec{y} - \beta_1 \cdot \vec{x}_1 - \beta_2 \cdot \vec{x}_2 \|_2^2.$$

$$\min_{\beta_1, \beta_2} \left\| \begin{pmatrix}1\\2\\3\end{pmatrix} - \beta_1 \begin{pmatrix}1\\0\\0\end{pmatrix} - \beta_2 \begin{pmatrix}0\\1\\0\end{pmatrix} \right\|_2^2$$

$$= \min_{\beta_1, \beta_2} \left((1-\beta_1)^2 + (2-\beta_2)^2 + (3-0)^2 \right).$$

$$\boxed{\beta_1^* = 1, \quad \beta_2^* = 2.}$$

$$X = \begin{bmatrix}1 & 0\\0 & 1\\0 & 0\end{bmatrix} \qquad Y = \begin{pmatrix}1\\2\\3\end{pmatrix}.$$

$$\boxed{\cancel{\beta^* = \frac{X^T y}{X^T X}}}$$

$$\beta^* = (X^T X)^{-1} \cdot X^T y.$$

$$\boxed{X^T y = \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\end{pmatrix} \begin{pmatrix}1\\2\\3\end{pmatrix}}$$

$$= \begin{pmatrix}1\\2\end{pmatrix}.$$

$$X^T X = \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\end{pmatrix} \begin{pmatrix}1 & 0\\0 & 1\\0 & 0\end{pmatrix}$$

$$\beta^* = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix} \begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}1\\2\end{pmatrix} \quad !!$$

$$\boxed{X^T X = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}}$$

What could wrong?
$X^T X$ could be non invertible. then
You have multiple solutions.

6) in Linear Regression you have features

x_1 x_2 x_3 y.

You try to Project the target column y on the feature columns x_1 x_2 x_3.