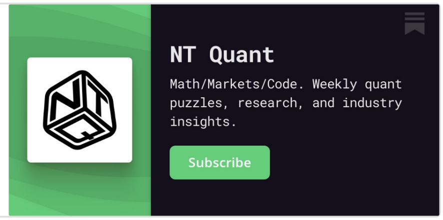
Quant Puzzles 6

NT Quant | Substack

Math/Markets/Code. Weekly quant puzzles, research, and industry insights. Click to read NT Quant, a Substack publication with hundreds of readers.

https://ntquant.substack.com/



Question 1 - Holding Spades [Easy]

4 people are separated into teams A and B comprised of 2 people each. A deck of cards is then dealt to the four players. Is it more likely that team A holds all the spades or none of the spades?

Solution:

both events are equally likely. Many people can solve this using pure intuition but here is a more formal proof.

Let X be the event that team A holds all the spades. Let Y be the event that team B holds all the spades. Let Z be the event that team A holds none of the spades. Let N be the event that team B holds none of the spades.

- If team A holds all the spades (X), then team B holds none of the spades **(N)** and vice versa. Therefore, the probability of both events is equal.

$$P(X) = P(N)$$

- If team B holds all the spades then team A holds none of the spades and vice versa

$$P(Y) = P(Z)$$

- Since the cards are dealt randomly, the probability that A holds all the spades and B holds all the spades are equal

$$P(X) = P(Y)$$

$$P(X) = P(Y) = P(Z)$$

Therefore, the probability that A holds all the spades is equal to the probability that A holds none of the spades.

Question 2 - Normal Equation [Medium]

- This question is a very basic machine learning question but might be hard if you haven't taken a course in advanced linear algebra or statistical learning. If this is the case, leave a comment on the post and I'll know whether I should write some ML Theory content. Additionally, if you have no idea how to approach this problem go ahead and look at the solution.

The cost function of fitting a linear regression is

$$RSS(\theta; X) = \sum_{i=1}^{N} (y^{(i)} - X^{(i)}\theta)^2$$

where X is the data matrix, the superscript is the row of the matrix and theta is a vector of the regression weights. Derive the weight vector that minimizes the value of the loss function.

Solution:

First, we convert the function into a matrix form

$$RSS(\theta; X) = \sum_{i=1}^{N} (y^{(i)} - X^{(i)}\theta)^2 = (\vec{y} - X\theta)^T(\vec{y} - X\theta)$$

$$= (\vec{y}^T - \theta^T X^T)(\vec{y} - X\theta) = \vec{y}^T\vec{y} - \vec{y}^T X\theta - \theta^T X^T \vec{y} + \theta^T X^T X\theta$$

The expression below evaluates to the exact same constant. Therefore we can simplify the equation as follows

$$\vec{y}^T X\theta = \theta^T X^T \vec{y}$$

$$\implies RSS(\theta; X) = \vec{y}^T\vec{y} - 2\theta^T X^T \vec{y} + \theta^T X^T X\theta$$

$$\nabla_\theta RSS(\theta; X) = -2X^T\vec{y} + 2X^T X\theta = 0$$

$$\implies X^T X\theta = X^T\vec{y}$$

$$\implies \theta = (X^T X)^{-1} X^T \vec{y}$$

In case the computation of the derivative seemed hard to understand, we will break it down in excruciating detail.

- The first expression doesn't contain any theta term so it evaluates to 0

$$\nabla_\theta \vec{y}^T\vec{y} = 0$$

- The next expression can be multiplied out to observe the result

$$-\nabla_\theta 2\theta^T X^T \vec{y} = \nabla_\theta \left(-2 \begin{bmatrix} \theta_1 & \dots & \theta_n \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{n1} \\ . & & . \\ . & & . \\ . & & . \\ x_{1m} & \dots & x_{nm} \end{bmatrix} \begin{bmatrix} y_1 \\ . \\ . \\ . \\ y_n \end{bmatrix}\right) = -2\nabla_\theta \begin{bmatrix} \theta_1 \sum_{i=1}^{n} x_{i1} y_i \\ . \\ . \\ . \\ \theta_n \sum_{i=1}^{n} x_{im} y_i \end{bmatrix} =$$

$$-2 \begin{bmatrix} \frac{d}{d\theta_1} \theta_1 \sum_{i=1}^{n} x_{i1} y_i \\ . \\ . \\ . \\ \frac{d}{d\theta_n} \theta_n \sum_{i=1}^{n} x_{im} y_i \end{bmatrix} =$$

$$-2\nabla_\theta \begin{bmatrix} \sum_{i=1}^{n} x_{i1} y_i \\ . \\ . \\ . \\ \sum_{i=1}^{n} x_{im} y_i \end{bmatrix} = -2X^T \vec{y}$$

- The third term is complicated to break down but if you multiply the matrix out, you get a sum of quadratic thetas, differentiating them will provide a linear term. and thus

$$\frac{d}{d\theta} \theta^T X^T X \theta = 2\theta^T X^T$$