01:XXX:XXX - Homework n

Pranav Tikkawar

April 7, 2025

JPD: Joint Probability Distribution

$$P(X = x, Y = y) = P(X = x)P(Y = y|X = x) = P(Y = y)P(X = x|Y = y)$$

Regression

$$Y|X \sim \mathcal{N}(\mu = x^T\Theta, \sigma^2 = \text{idp. of x})$$

Suppose we have data tuple $\{(x_i, y_i)\}_{i=1}^n$ Which are observations from indept RV $\{Y|x_i\}$ Now we know that these data appear, what can be said about these parameters.
Maximum Likelihood Estimation: Maximize the likelihood of the data given the parameters.

$$\Theta_{MLE} = \text{argmax}_\Theta \prod_{i=1}^{n} P(Y = y_i|X = x_i)$$

We have $J(\theta)$ to meausre the accurate the model is.

$$J(\theta) = \sum_{i=1}^{n}(y_i - x_i^T\theta)^2$$

for the MLE approach Consider $J(\theta + h)$

$$\begin{aligned} J(\theta + h) &= ||y - X(\theta + h)||^2 \\ &= J(\theta) + \nabla J(\theta)h + \frac{1}{2}h^T\nabla^2 J(\theta)h + o(||h||^3) \end{aligned}$$

1 Principal Component Analysis

Problem: "Reduce dimension/compress data/ fewer numbers"
Naturally if we have like 4 points in $\mathbb{R}^2$ like $(1,5),(2,7),(3,2),(5,5)$ a natural choice to reduce numbers is just to take x or just y coordinates.
Instead we can use the idea of disance can cauluate the distance between the points.
We want to make the 4 numbers as seperated as possible.
for all points x_i then we can take the mean of the points and then take the distance from the mean.

$$\mu = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2$$

$$= 2(\frac{1}{N}\sum_i x_i^2 - \left(\frac{1}{n}\sum_i x_i\right)^2)$$

We know the empirical mean of data is $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
To measure the spread we take the empirical variance of the data.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

So if empirical mean is 0 then we are left with $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$
We want to find such that the variance is maximized.

$$V[z] = \sum_{i=1}^{n} ||z_i||^2$$

We will look for a linear encoding, ie the codes are obtained by applying a lear map to the input

$$z_n = B^T x_n$$

Where B is a $D \times M$ matrix. with $z \in R^M$ and $x \in R^D$
B has m columns in R^D
Now our multiplicaiton is like an inner product.
IE the ith component of z_n is $b_i \cdot x_n$
Now if we take our columsn $b_1 ... b_m$ to be orthonormal then we can write the variance as

$$V[z] = \sum_{i=1}^{m}\sum_{n=1}^{N}(b_i \cdot x_n)^2$$

Or we need to max $V \sum b^t x_n x_n^t b$
Eventually with some fun math we get to $b_1^t s b_1$
Find b_1 st $V_1 = b_1^t S b_1$ is maximized among unit vectors

$$b_1^t S b_1 - 1 = 0$$

Thus we solve lagrange multiples

$$\frac{d}{db_1}(b_1^t S b_1) = \lambda \frac{d}{db_1}(b_1^t b_1 - 1)$$