1

Class 8 homework

First - to revisit problem 1 from the $\boxed{\text{class 7}}$ homework:

Solve for β_0, β_1 by minimizing MSE:

Substitute $\hat{y}_i = \beta_0 + \beta_1 x_i$ into $\sum (y_i - \hat{y}_i)^2$

To minimize $\sum (y_i - \beta_0 - \beta_1 x_i)^2$, we can show $\beta_0 \to 0$ when the ys & xs are centralized:
Start by solving for β_0

$$0 = \frac{\partial \sum (y_i - \beta_0 - \beta_1 x_i)^2}{\partial \beta_0} = 2 \sum (y_i - \beta_0 - \beta_1 x_i)(1)$$
$$= 2 \sum y_i - n\beta_0 - \beta_1 \sum x_i$$

Or:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

This means that β_0 is just a function of the means of y & x.
So if y & x are centered $\beta_0 \to 0$.

This makes the solution of β_1 simpler:

Let $x' = x - \bar{x}$, $y' = y - \bar{y}$, $\beta_0 = 0$

$$0 = \frac{\partial}{\partial \beta_1} \sum (y_i' - \beta_1 x_i')^2 = \sum 2(y_i' - \beta_1 x_i') \frac{\partial}{\partial \beta_1}(y_i' - \beta_1 x_i')$$
$$= \sum y_i' x_i' - \beta_1 \sum x_i' x_i'$$

So $\beta_1 = \dfrac{\langle y' \cdot x' \rangle}{\langle x' \cdot x' \rangle}$

If we substitute x & y into the β_1 we get

$$\beta_1 = \frac{\mathrm{cov}(y, x)}{\mathrm{var}(x, x)}$$

Class 8 HW Solutions

Entropy

i) Use $MI(P;Q) = H(P) - H(P|Q)$

To show

$$MI(P;Q) = -\sum_{ij} P(x_i\, y_i)\log\left[\frac{P(x_i)P(y_i)}{P(x_i\, y_i)}\right]$$

$$MI(P;Q) = E_P\left[\log P(x_i)\right] - E_{PQ}\left[\log\left(P(x_i | y_i)\right)\right]$$

$$= E_{PQ}\left[\log P(x_i) - \log\left(\frac{P(x_i\, y_i)}{P(y_i)}\right)\right]$$

$$= E_{PQ}\left[-\log\left(\frac{P(x_i\, y_i)}{P(x_i)P(y_i)}\right)\right]$$

$$= -\sum_{ij} P(x_i\, y_j)\log\left(\frac{P(x_i)P(y_j)}{P(x_i\, y_j)}\right)$$

Class 8 Solutions

Linear Algebra

1) To show matrices w/ the same eigenvectors commute, note their similarity matrices are the same:

$$A = S^{-1}\Lambda_A S \;,\quad B = S^{-1}\Lambda_B S$$

$$AB = S^{-1}\Lambda_A S S^{-1}\Lambda_B S$$

$$= S^{-1}\Lambda_B \Lambda_A S$$ Since diagonal matrices commute

$$= S^{-1}\Lambda_B S S^{-1}\Lambda_A S = BA.$$

2. A matrix with a non-zero nullspace implies there are $x \neq 0$ st $Ax = 0$
so $(A-\lambda I)x = Ax = 0$ if $\lambda = 0$

3. Adding duplicate rows to the design matrix doesn't change the span of its column space - e.g. if the columns were lin. indep. they remain so.

So - duplicating all the rows has no effect on the regression results.

Adding a selected set of rows gives them a higher weight - as if they are more prevalent in the population.