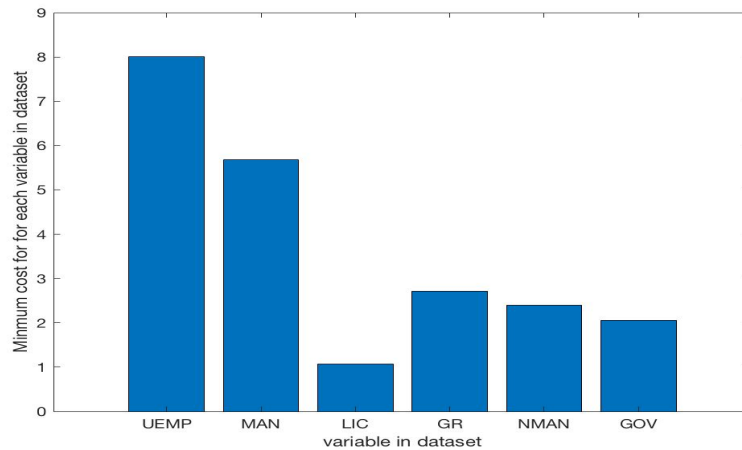


Report CS542

Shifeng Li U24567277

in program a, I use linear regression to compute these data. At first, I should use the matrix to express the data, and choose one column to become the third variable. Then I test any column to find the best, by compute the cost function.



From the photo, we can find that when we choose the column of LIC the cost is minimize.

the linear combination of the input

$$y(x, w) = w_0 + \sum_{j=0}^{M-1} w_j \phi_j(x) \quad (\text{Bioshop 3.2})$$

basis function :

$$y(x, w) = w_0 + \sum_{j=0}^{M-1} w_j \phi_j(x) = w^T \phi(x) \quad (\text{Bioshop 3.3})$$

In the basis function, I set $\phi(x) = x$, Then consider a total error function which combines Bioshop(3.12) and the regularization (3.25) . Calculate the error for each input x with its w_{ML} and choose the input x with minimum error as the best input. The minimum error means the result we get is the closest to the actual value

2 KNN

if the data in a column is numerical, I use the mean of all the data in that

column to impute the missing value. If the data in a column is not numerical, I use the mode to impute the missing value.

In my program, the file and program should put into a same folder. Then running the program.

The accuracy for two different K of two datasets

K	crx	lenses
1	0. 7029	0. 8333
3	0. 7101	0. 6667
5	0. 7246	0. 6667