Regression: a basic stat tool

Stat 154, Fa13

N. El Karoui

Prostate cancer data

- Regression output:
- Coefficients:
- Estimate Std. Error t value Pr(>|t|)
- (Intercept) -6.718e-17 7.205e-02 0.000 1.00000
- lcavol 5.931e-01 1.105e-01 5.366 1.47e-06 ***
- lweight 2.423e-01 8.808e-02 2.751 0.00792 **
- age -1.180e-01 8.455e-02 -1.396 0.16806
- lbph 1.755e-01 8.538e-02 2.056 0.04431 *
- svi 2.563e-01 1.038e-01 2.469 0.01651 *
- lcp -2.393e-01 1.282e-01 -1.867 0.06697 .
- gleason -1.732e-02 1.180e-01 -0.147 0.88389
- pgg45 2.296e-01 1.321e-01 1.738 0.08755

- What does it mean? What does it implicitly assume?

Least squares

- Call the response vector Y and the matrix of predictors X
- Often times, first column of X is full of 1s
- We have

$$RSS(\beta) = \|Y - X\beta\|_2^2$$

- This is the matrix form of the usual least-squares minimization problem.
- Solution?

LS Solution

- After elementary manipulations seen on the board, if $X'X$ is invertible, we have

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y$$

- Fitted values:

$$\hat{Y} = X\hat{\beta} = X\left(X'X\right)^{-1} X'Y = HY$$

Various issues

- p>n?

$$\sigma^2$$

Various issues

- p>n?
- Estimate of σ^2 ?

Various issues

- p>n?
- Estimate of σ^2 ?
- How to derive CI's?

Various issues

- p>n?
- Estimate of σ^2 ?
- How to derive CI's?
- The problem of correlated predictors
- Impact on CI's/prediction?

Various issues

- p>n?
- Estimate of σ^2 ?
- How to derive CI's?
- The problem of correlated predictors
- Impact on CI's/prediction?
- Geometric interpretation

Various issues

- p>n?
- Estimate of σ^2 ?
- How to derive CI's?
- The problem of correlated predictors
- Impact on CI's/prediction?
- Geometric interpretation
- Gauss-Markov and BLUE

Various issues

- p>n?
- Estimate of σ^2 ?
- How to derive CI's?
- The problem of correlated predictors
- Impact on CI's/prediction?
- Geometric interpretation
- Gauss-Markov and BLUE
- Leverage points

Various issues

- p>n?
- Estimate of σ^2?
- How to derive CI's?
- The problem of correlated predictors
- Impact on CI's/prediction?
- Geometric interpretation
- Gauss-Markov and BLUE
- Leverage points
- QR Decomposition and consequences

Subset Selection

- More predictors do not always help
- Statistically, may degrade prediction accuracy. Example?
- Issue of interpretability.
- We will be very interested in subset/model selection methods

Subset selection

- Idea: try subsets of various size
- Find the one that gives the best prediction error
- Issue?

Subset selection

- Idea: try subsets of various size
- Find the one that gives the best prediction error
- Issue?
- Often, in practice, best subset is picking by cross-validation
- Rule often used: use most parsimonious subset within 1 sd of best EPE

Forward-stepwise

- Addresses computational problem of best subset
- Idea: suppose current model has k variables
- Add 1 extra variable to the model, chosen among remaining p-k
- Pick variable that improves fit most
- Now we have model with k+1 variables
- Iterate

Backward-stepwise

- Same idea as forward stepwise
- But move backwards from full model
- Why?

Backward-stepwise

- Same idea as forward stepwise
- But move backwards from full model
- Why?
- Concern about greediness of algo.
- Limitation: if p>n, can't do it…

Forward-stagewise

- Same rough idea as forward-stepwise, but takes more cautious steps…
- Given a current model, search in predictors variable that is most correlated with current residuals
- Regress that variable (e.g the k-th one) on the residuals. Call $\hat{\beta}_{k,M}$ the value of that coeff (M is for M-th step)
- Update $\hat{\beta}_k \leftarrow \hat{\beta}_{k,M} + \hat{\beta}_k$