What is an unbiased estimator

What is meant by best

Regularized regression Ridge, LASSO, Elastic Net

Compare and contrast Cross-validation

* Linear regression is the **best linear unbiased estimator** (BLUE)

if on avg is aiming towards the true , it is unbiased

* + notion that this beta hat this you have trying to estimate true beta , which you never see
  + If the distribution of statistic is centered over truth - its called unbiased

Best

* + IF assumptions hold, is the minimum variance of all unbiased estimators
  + **If assumptions hold**, the spread of s (every sample gives a different sampling distribution spread) won’t be overly wide
  + Best - spread of guesses are as narrow as it get

***Best, unbiased***- If it is aiming at the middle of the right target, the spread isn't wide

As the number of variables increases

* + Assumptions start to fail, multicollinearity concerns
  + Overfitting

**Regularized Regression**

penalizing the model, changing the coefficients by shrinking them

* + Giving up interpretability of for the hope of being able to predict better
* Coefficients become biased, but potentially improve variance of the model
  + Changing coefficient and makes it biased (no longer aiming for true relationship)

Willing to give up interpretability to relax assumptions to make prediction better

* + Regularized regression - can have multicollinearity

Penalties

* Ordinary least squares (OLS) minimizes the sum of squared errors (SSE)
* Regularized regression introduces a penalty term to the minimization
* **Ridge**- introduces an “L2” penalty term to the minimization:, alpha =0
  + if , the its OLS (there is no penalty)
  + 𝜆 🡪 ∞, coefficients shrink to 0
  + **LASSO**- Least absolute shrinkage and selection operator (LASSO) regression ,
  + alpha = 1
  + introduces an “L1” 𝜆 🡪 ∞, coefficients shrink to 0
    - Ridge: approaches 0 asymptotically, lasso can remove variables

**Elastic NET**

* + LASSO does variable selection ; Ridge keeps all variables (LASSO drops arbitrarily
  + Elastic net regression combines both penalty terms in the minimization
  + Any value of α between 0 and 1 gives a combination of both penalties (elastic net)
  + They can be zeroed out but at a different way that of LASSO because we have ridge part as well.

Optimizing penalty:

Fear of Overfitting

* Need to select 𝜆 for any of the regularized regression approaches
* Don’t want to minimize variance to the point of overfitting our model to the training data

**CROSS VALIDATION**: Cross-validation (CV) 🡪 approach to prevent overfitting when tuning a parameter

1. Split training data into multiple pieces
2. Build model on majority of pieces
3. Evaluate on remaining piece
4. Repeat process with switching out pieces for building and evaluation

Minimum MSE, find beta that minimizes the error // mathematically best uses all vars

One standard error above the minimum 🡺 if willing to account for variability pf the dots, you can get something close to the minimum error

Can get rid of vars to get close to minimum 0, one SE above to get the close error so model can have less var (trade off for min error)

one standard error above the minimum value. This value is especially useful in LASSO regressions. The largest λ� within one standard error would provide approximately the same MSE, but with a further reduction in the number of variables.

A diagram of mathematical equations

Description automatically generated

LASSO

1. Selecting variables
2. Converting categorical variables to factors
3. Separating target variable and predictor variables
4. Splitting into train and test data
5. model.matrix() - for creating dummy codes for categorical variables
6. Isolate the target variable in its vector
7. Repeat the above two steps for test data as well.
8. Modelling cv.glmnet() - obtain lambda values
9. Investigate coefficients for min and max lambda
10. Pick one lambda
11. Get predictions for test data
12. check the model metrics