INFO251 – Applied Machine Learning

Lab 6

Suraj R. Nair & satej soman

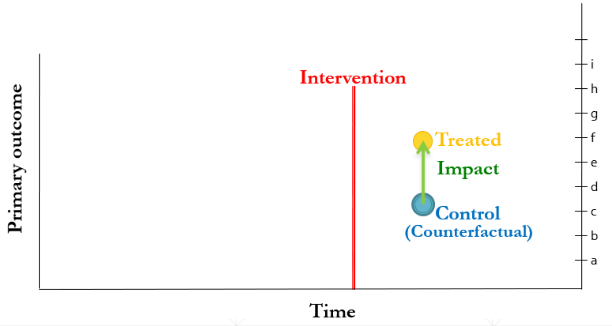
Announcements

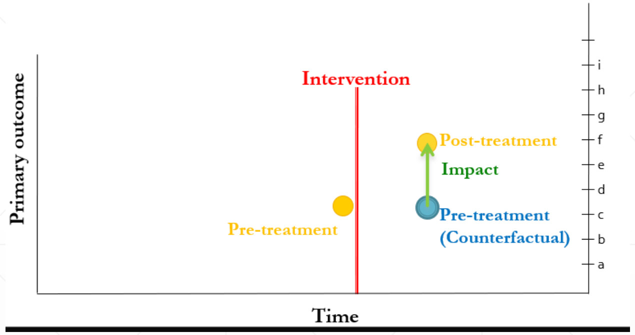
- Quiz 1 on March 4
- PS2 Grades released, PS3 Grades next week
- PS 4 due on March 13
- Today:
  - Quiz review: code demo + quiz questions discussion
  - For derivations / discussions related to mathematical intuition: office hours

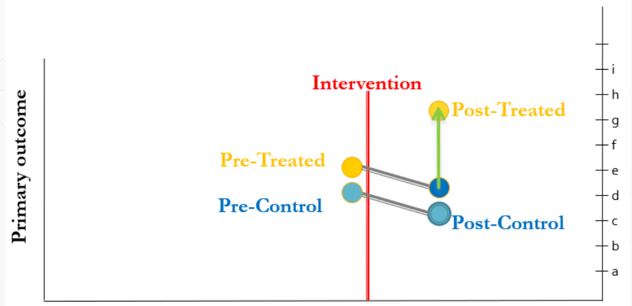
Quiz 1 Review

Quick Review: Research Designs

Design	Key Identifying Assumption	Confounds / Threats to identification
Randomized experiment (T v/s C)	?	?
Pre v/s Post	?	?
Double Difference	?	?







Quick Review: Research Designs

Design	Key Identifying Assumption	Confounds?
Randomized experiment (T v/s C)	*Outcomes* in Treatment and Control *would have* been the same in the absence of treatment	Differential selection / non-compliance / attrition
Pre v/s Post	*Outcomes* pre- and post-treatment *would have* been the same in the absence of treatment	Temporal Trends, Seasonality, etc.
Double Difference	In the absence of treatment, trends in the outcome variable for the treatment group and the control group would have been the same	Differential trends

1. Diff-in-diff

- Suppose you are evaluating the impact of a minimum wage program on employment rates. In the treatment group, the employment rate changed from 74% (pre) to 82% (post). In the control group, during the same time, the employment rate changed from 71% (pre) to 68% (post).
- Estimate the true impact of the minimum wage program.

1. Diff-in-diff: Solution

(82 – 74) – (68 - 71) = 8 + 3 = 11

2. Linear Regression

We run a linear regression of the form

$$GPA = \alpha + \beta \text{ StudyingHours} + \gamma \text{ ChatGPT}$$

StudyingHours is continuous (time spent reviewing lecture notes); ChatGPT is binary (indicator for whether student uses ChatGPT to write assignments)

$\alpha = 0.5$

$\beta = 0.12$

$\gamma = -0.05$

What is the difference between the GPA of a student who spends 20 hours studying + uses ChatGPT, and the GPA of a student who spends 40 hours studying + does not use ChatGPT?

2. Linear Regression: Solution

- 20hrs + ChatGPT = 0.5 + 0.12 * 20 – 0.05
- 40hrs + No ChatGPT = 0.5 + 0.12 * 40
- Difference in GPA = 2.45

3. Logistic Regression

- $\text{logit}(honor_i) = \alpha + \beta STEM_i + \epsilon_i$
- Calculate (from the regression results below):
  - odds of a non-STEM student pursing an honors degree?
  - odds of a STEM student pursuing an honors degree?
  - the odds ratio (STEM vs Non-STEM)
  - probability that a STEM student is an honors student?

hon	stem: no	stem: yes	Total
0	74	77	151
1	17	32	49
Total	91	109	200

```
Logistic regression                             Number of obs     =        200
                                                LR chi2(1)        =       3.10
                                                Prob > chi2       =     0.0781
Log likelihood = -109.80312                     Pseudo R2         =     0.0139
```

hon	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
stem	.5927822	.3414294	1.74	0.083	-.0764072	1.261972
intercept	-1.470852	.2689555	-5.47	0.000	-1.997995	-.9437087

3.Logistic Regression: Solution

- odds of a non-STEM student pursing an honors degree: exp(-1.470852)
- odds of a STEM student pursuing an honors degree?: exp(-1.470852 + 0.5927822)
- probability that a STEM student is an honors student?:
  - log(p / 1-p) = -1.470852 + 0.5927822
  - p = 1 / (1 + 1/exp(-1.470852 + 0.5927822)) = 0.2936

3.Logistic Regression: Solution (Derivations)

This derivation shows that β from the previous slide gives us the log odds ratio.

This is our logistic regression:

log(odds (Y = 1 | STEM)) = log $\left(\frac{P(Y=1\,|STEM)}{1-P(Y=1\,|\,STEM)} \right) = \alpha + \beta\ STEM_i$

Compare the log odds for STEM = 0, and STEM = 1

log(odds (Y = 1 | STEM = 1)) - log(odds (Y = 1 | STEM = 0)) $= \alpha + \beta\ - \alpha = \beta$

→ log $\left(\frac{\text{odds (Y = 1 | STEM = 1)}}{\text{odds (Y = 1 | STEM = 0)}} \right) = \beta$ [using basic properties of log: log(a) – log(b) = log(a/b)]

→ $\underbrace{\frac{\text{odds (Y = 1 | STEM = 1)}}{\text{odds (Y = 1 | STEM = 0)}}}_{\text{Odds ratio}} = e^{\beta}$ [exponentiating, and getting rid log on the LHS]

3.Logistic Regression: Solution (Derivations)

- e^{β} = exp(0.5927822) = 1.809
- From the cross-tabs, we see that the odds ratio is:

hon	stem: no	stem: yes	Total
0	74	77	151
1	17	32	49
Total	91	109	200

  - odds (Y = 1 | STEM = 1) = 32 / 77 ≈ 0.42
  - odds (Y = 1 | STEM = 0) = 17 / 74 ≈ 0.23
  - odds ratio = $\frac{\text{odds (Y = 1 | STEM = 1)}}{\text{odds (Y = 1 | STEM = 0)}}$ ≈ 0.42 / 0.23 ≈1.809
  - Interpretation: the odds for a STEM student to be enrolled in an Honors program are 81% higher!

4. Ridge regression

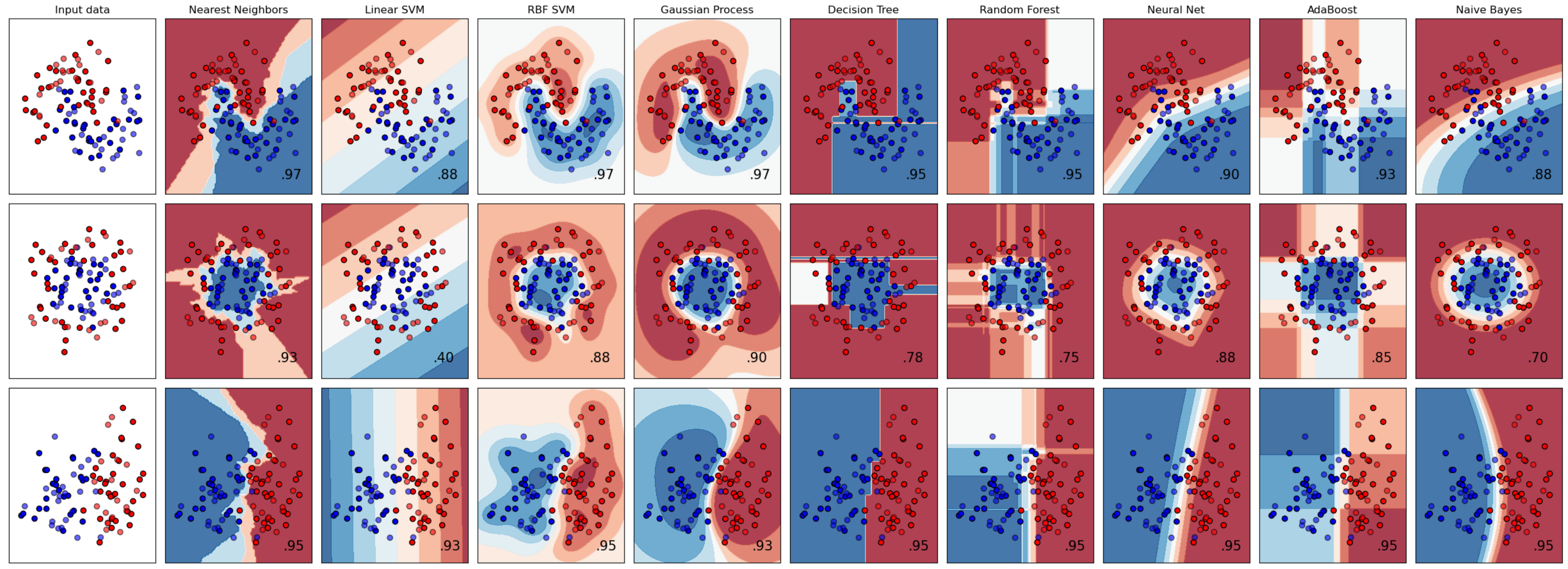
- Statement A: As the regularization penalty becomes larger, ridge regression coefficients approach infinity
- Statement B: Ridge regression forces some coefficients to zero

1. A is True, B is True
2. A is True, B is False
3. A is False, B is True
4. A is False, B is False

4. Ridge Regression: Solution

- A is False, B is False

Quick Review: Decision Boundaries



[Source]

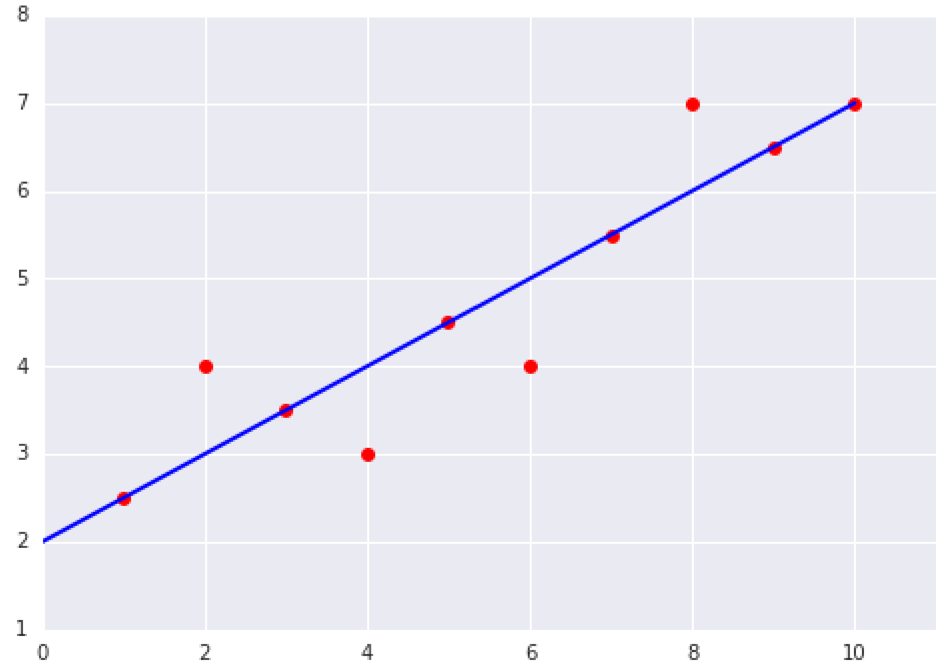
5. Decision Boundaries

- Which of the following algorithms recovers non-linear decision boundaries:
  - K-nearest neighbors (K = 5)
  - SVM
  - Logistic Regression
  - Logistic Regression with lasso regularization

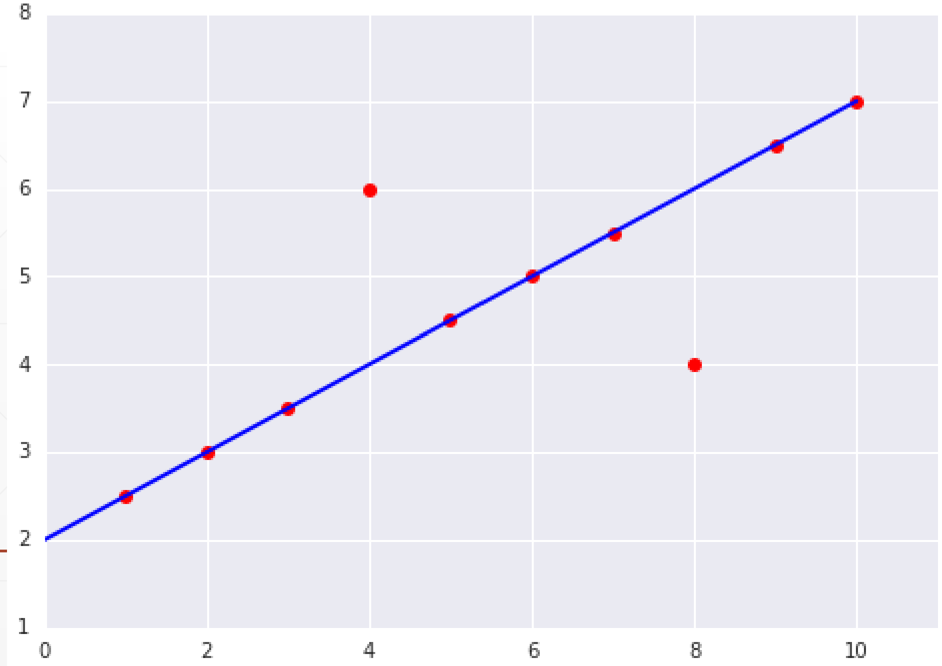
6. Mean Squared Error

- Suppose you build a linear regression model which predicts y = f(x). Which of these two cases has a higher MSE?
  - A
  - B

A



B



6. MSE Solution

- MSE for A: 4/ 10
- MSE for B: 8 / 10 → higher MSE

7. Bayes Theorem

- A doctor knows that having a cold causes you to sneeze 50% of the time.
- Prior probability of any patient having a cold is 1/10,000
- Prior probability of any patient sneezing is 1/15
- If a patient is sneezing, what is the probability they have a cold?

7. Bayes Theorem: Solution

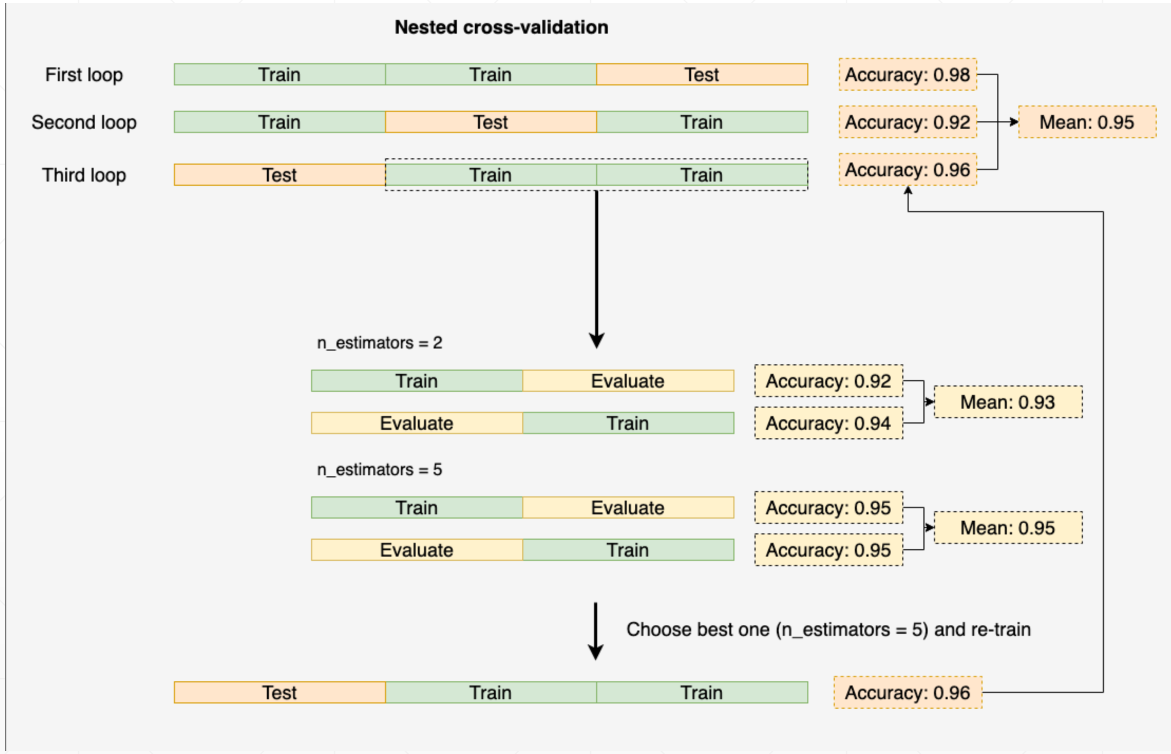
- Let S = sneezing, C = Cold
- $P(S \mid C) = 0.5$ (from “a doctor knows that having a cold causes you to sneeze 50% of the time”)
- $P(C) = 1/10000$ (Prior probability of any patient having a cold is 1/10,000)
- $P(S) = 1/15$ (Prior probability of any patient sneezing is 1/15)
- We need to find $P(C \mid S)$ (If a patient is sneezing, what is the probability they have a cold)
- From Bayes Rule:
  - $P(C \mid S) = P(S \mid C) * P(C) / P(S)$
  - $P(C \mid S) = (0.5 * 1/10000) / (1/15) = 0.00075$

8. Cross-validation

- Suppose you want to estimate the out of sample performance of a K-nearest neighbors algorithm using nested cross-validation. If you have 5 outer loops, 10 inner loops and 20 different values for K in the hyperparameter grid, how many times will the learning algorithm nearest_neighbor(K) be called?
- *Hint: Don't forget the refit step!*

8. Cross Validation: Solution

- Each inner loop: 10 * 20 function calls
- This happens 5 times ---> 5 * (10 * 20) calls
- Each outer loop includes 1 refit call ---> 5 * (10 * 20) + 5 = 1005



9. Classification

- Calculate accuracy, TPR, FPR and Precision for the “green” class.

		Predicted	
		Green	**Orange**
Actual	**Green**	9	3
	Orange	2	1

9. Classification: Solution

- For the “green” class:

		Predicted	
		Green	**Orange**
Actual	**Green**	True Positives	False Negatives
	Orange	False Positives	True Negatives

		Predicted	
		Green	**Orange**
Actual	**Green**	9	3
	Orange	2	1

- Accuracy: (TP + TN) / (TP + FP + FN + TN) = 10 / 15
- TPR (or Recall) for green class= TP / (TP + FN) = 9/12
- FPR for green class = FP / (FP + TN) = 2 / 3
- Precision for green class = TP / (TP + FP) = 9 / 11