Data100 Sp22 Disc 8
OHE/Regularization

Attendance:

https://tinyurl.com/disc8michelle

Announcements

Due Dates

- Lab 8 due Tues, March 15
- Proj 1B (Housing) due Thurs, March 17

Other

-We will be starting exam prep sessions!

Fairness in Housing Appraisal

1

Fairness in Housing Appraisal

Recall that Project 1's dataset comes from the Cook County Assessor's Office (CCAO) in Illinois, a government institution that determines property taxes across most of Chicago's metropolitan area and its nearby suburbs. In the United States, all property owners are required to pay property taxes, which are then used to fund public services including education, road maintenance, and sanitation.

1. "How much is a house worth?" If you were a homeowner, why would you want your property to be valued high? Why would you want your property to be valued low?

2. Which of the following scenarios strike you as unfair and why? You can choose more than one. There is no single right answer but you must explain your reasoning.

   A. A homeowner whose home is assessed at a higher price than it would sell for.
   B. A homeowner whose home is assessed at a lower price than it would sell for.
   C. An assessment process that systematically overvalues inexpensive properties and undervalues expensive properties.
   D. An assessment process that systematically undervalues inexpensive properties and overvalues expensive properties.

3. Imagine your home is assessed at a higher value than you believe it would sell for on the market. What might that concretely mean to you, as an individual homeowner?

Answer list

Q1

- Sell your house - want a high valuation
- Taxes (minimize burden during assessment) - want a low valuation

Q2

A. **Unfair on homeowner -** pays more taxes
   Unfair on buyer - more difficult to buy a home if overpriced
B. **Unfair on community -** pays lower taxes than they should
   Unfair on seller - sold for lesser than valuation
C. Putting greater tax burden on lower socio-economic status (consistently overvalued)
D. Widens wealth gap - hard for middle class to move up/purchase property

One Hot Encoding

2

One Hot Encoding Motivation

- Encoding “dummy” variables
- Useful for categorical variables

id	color
1	red
2	blue
3	green
4	blue

One Hot Encoding →

id	color_red	color_blue	color_green
1	1	0	0
2	0	1	0
3	0	0	1
4	0	1	0

In order to include a qualitative variable in a model, we convert it into a collection of dummy variables. These dummy variables take on only the values 0 and 1. For example, suppose we have a qualitative variable with 3 possible values, call them A, B, and C, respectively. For concreteness, we use a specific example with 10 observations:

$$[A, A, A, A, B, B, B, C, C, C]$$

We can represent this qualitative variable with 3 dummy variables that take on values 1 or 0 depending on the value of this qualitative variable. Specifically, the values of these 3 dummy variables for this dataset are x_A, x_B, and x_C, arranged from left to right in the following design matrix, where we use the following indicator variable:

$$x_{k,i} = \begin{cases} 1 & \text{if } i\text{-th observation has value } k \\ 0 & \text{otherwise.} \end{cases}$$

This representation is also called one-hot encoding. It should be noted here that $\vec{x_A}$, $\vec{x_B}$, and $\vec{x_C}$ are all vectors.

$$\mathbb{X} = \begin{bmatrix} | & | & | \\ \vec{x_A} & \vec{x_B} & \vec{x_C} \\ | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

We will show that the fitted coefficients for $\vec{x_A}$, $\vec{x_B}$, and $\vec{x_C}$ are $\bar{y}_A$, $\bar{y}_B$, and $\bar{y}_C$, the average of the y_i values for each of the groups, respectively.

4. Show that the columns of $\mathbb{X}$ are orthogonal, (i.e., the dot product between any pair of column vectors is 0).

5. Show that

$$\mathbb{X}^T\mathbb{X} = \begin{bmatrix} n_A & 0 & 0 \\ 0 & n_B & 0 \\ 0 & 0 & n_C \end{bmatrix}$$

Here, n_A, n_B, n_C are the number of observations in each of the three groups defined by the levels of the qualitative variable.

6. Show that

$$\mathbb{X}^T\mathbb{Y} = \begin{bmatrix} \sum_{i\in A} y_i \\ \sum_{i\in B} y_i \\ \sum_{i\in C} y_i \end{bmatrix}$$

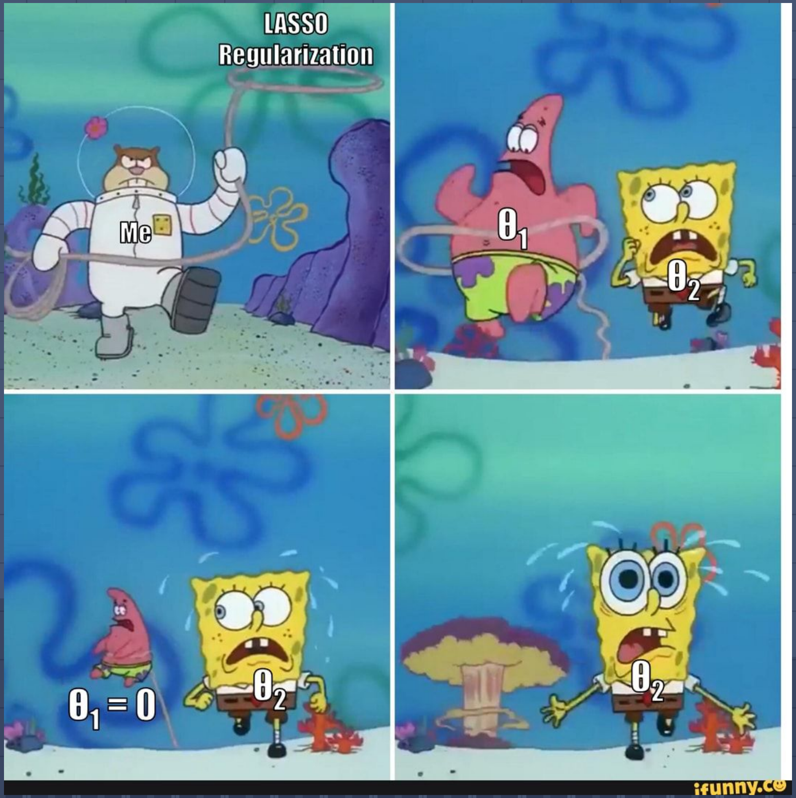
where i is an element in group A, B, or C.

7. Use the results from the previous questions to solve the normal equations for $\hat{\theta}$, i.e.,

$$\begin{aligned} \hat{\theta} &= [\mathbb{X}^T\mathbb{X}]^{-1}\mathbb{X}^T\mathbb{Y} \\ &= \begin{bmatrix} \bar{y}_A \\ \bar{y}_B \\ \bar{y}_C \end{bmatrix} \end{aligned}$$

Regularization

3



LASSO Regularization
Me
θ1
θ2
θ1 = 0
θ2
θ2
ifunny.co

Regularization

-Help us prevent overfitting
-Add a “cost” to to having large thetas

$$\frac{1}{n}\sum_{i=1}^{n} Loss(y_i,\ \hat{y_i}) + \lambda R(\theta)$$

Ridge (L2) vs. LASSO (L1)

-Remember our loss functions! Analogous to L1/L2 loss
-Main differences:
- L1 is **sparser**, shrinks the less important features coefficients to zero, thus removing the feature all together
- L1 has no closed form solution, L2 does

LASSO Regression:

$$\frac{1}{n}||Y - X\theta||_2^2 + \lambda \sum_{j=1}^{d} |\theta_i|$$

Ridge Regression:

$$\frac{1}{n}||Y - X\theta||_2^2 + \lambda \sum_{j=1}^{d} \theta_i^2$$

(a) As model complexity increases, what happens to the bias and variance of the model?

out of scope

(b) In ridge regression, what happens if we set $\lambda = 0$? What happens as λ approaches ∞?

$\lambda = 0 \Rightarrow$ End up with OLS (original Equation)

$\lambda = \infty \Rightarrow \theta$ tends toward 0

(c) If we have a large number of features (10,000+) and we suspect that only a handful of features are useful, which type of regression (Lasso vs Ridge) would be more helpful in interpreting useful features?

LASSO sets many coefficients to zero, select most useful features.

(d) What are the benefits of using ridge regression over OLS?

* - when we have many useful features (not LASSO)

If not full rank infinite solutions

- $\hat{\theta} = (X^T X + n\lambda I)^{-1} X^T Y$ is closed form solution for ridge regression