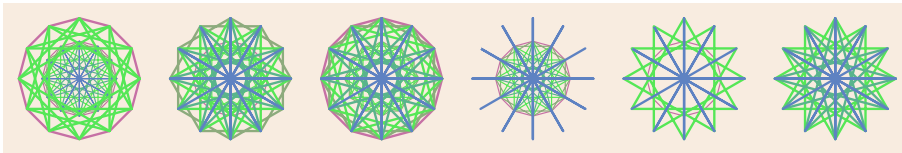


Quiz 1

CSC316 Machine Learning
Professor Leon Tabak

18 February 2022

This work is licensed under CC BY 4.0. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.



Questions

1. We prepared the California Housing Prices data before building our linear regression model.

List the several kinds of changes that we made to the data.

-
- Replace missing values in the *total_bedrooms* column with the median value.
 - Replace *ocean_proximity*, a multivalued categorical variable with five possible values, with five binary categorical variables using a one hot encoding function.
 - Scale all numerical variables, so that the mean value is zero and the standard deviation is one in all columns with numerical values.

- Construct three new variables (*rooms_per_household*, *population_per_household*, *bedrooms_per_room*). Assign values to these new columns by combining values in other columns.

2. In a sentence or two, name and define this statistic.

$$\frac{TP}{TP + FP}$$

$$\frac{TP}{TP + FP} = \text{precision}$$

TP = the number of True Positives

FP = the number of False Positives

$TP + FP$ = the total number of positives

The Precision is the fraction of all positive reports that are correct reports.

3. In a sentence or two, name and define this statistic.

$$\frac{TP}{TP + FN}$$

$$\frac{TP}{TP + FN} = \text{recall}$$

TP = the number of True Positives

FN = the number of False Negatives

$TP + FN$ = the total number of positives

The Recall is the fraction of positive instances that the system correctly identifies.

4. The F1 statistic is a harmonic mean. The arithmetic mean of 2 and 10 is...

$$\begin{aligned}\text{arithmetic_mean}(2, 10) &= \frac{2 + 10}{2} \\ &= 6\end{aligned}$$

What is the harmonic mean of 2 and 10?

$$\begin{aligned}\text{harmonic_mean}(2, 10) &= \frac{2}{\frac{1}{2} + \frac{1}{10}} \\ &= \frac{2}{\frac{6}{10}} \\ &= \frac{20}{6} \\ &= 3 \frac{1}{3}\end{aligned}$$

5. We label the rows of a confusion matrix with the actual values of our samples.

We label the columns of a confusion matrix with the predicted values of our samples.

Let's suppose that we are looking at an element of the confusion matrix that lies in the row labeled 1 and the column labeled 7. At that row and column, we see the number 381.

That means the model looked at 381 samples that were really 1's and incorrectly classified them as 7's.

If we succeeded in building a model that correctly classified all samples, what would the confusion matrix look like?

In that case, the only non-zero values in the confusion matrix would appear on the main diagonal (the line that extends from the upper left corner to the lower right corner).

6. We do not have a good model if the ROC curve lies everywhere beneath the diagonal line that extends from (0.0, 0.0) to (1.0, 1.0) on the graph. Explain.

The ROC curve will be under the diagonal only if the number of false positives exceeds the number true positives.

7. What is the AUC on an ROC plot? What does the ROC curve look like when AUC approaches its greatest possible value?

AUC is “Area Under the Curve.”

In the best cases, the curve goes nearly straight up and then makes a hard turn to right near the top of the scale. The area under the curve is nearly 1.0.

8. A linear regression model predict a value by computing the dot product of a parameter vector with a feature vector.

$$\Theta = [\Theta_0, \Theta_1, \Theta_2]$$

$$\mathbf{x} = [1, x_1, x_2]$$

$$\Theta \cdot \mathbf{x} = \Theta \mathbf{x}^T$$

$$= [\Theta_0 \quad \Theta_1 \quad \Theta_2] \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$$

$$=?$$

$$\Theta \cdot \mathbf{x} = \Theta_0 + \Theta_1 \mathbf{x}_1 + \Theta_2 \mathbf{x}_2$$

9. The Normal Equation includes the transpose of a matrix. Show that you know what the transpose of a matrix is by writing the transpose of this 4×3 matrix:

$$\begin{bmatrix} S & U & N \\ S & K & Y \\ S & E & A \\ S & O & D \end{bmatrix}$$

$$\begin{bmatrix} S & S & S & S \\ U & K & E & O \\ N & Y & A & D \end{bmatrix}$$

10. The Normal Equation includes the inverse of a matrix. That matrix is the product of \mathbf{X}^T and \mathbf{X} .

The number of rows in \mathbf{X} is the number instances in our dataset. The number of columns in \mathbf{X} is the number of features in our dataset.

$\mathbf{X}^T \mathbf{X}$ is a square matrix.

How many rows and columns are in that square matrix?

Is it the number of instances in the dataset or is it the number of features in the dataset?

The number of rows and columns in the square matrix equals the number of columns in \mathbf{X} —this is the number of features in the dataset.

11. Stochastic Gradient Descent (SGD) computes $\nabla_{\Theta} \text{MSE}(\Theta)$ at each step. MSE is a convex function. Explain why this is good news.

There are no local minima or maxima, so the algorithm cannot get stuck in a local minimum. The algorithm will converge on a solution.

12. In a paragraph, summarize what you learned by reading in our book about “no free lunch” and the “unreasonable effectiveness of data.”

We cannot know which kind of model will give us the best results except by experimenting with different models. Because we cannot try every possible model, we must make some assumptions. We also know that, given enough data, many different models will produce very similar results. For these reasons, we might want to give priority to efforts to collect more and better data over efforts to improve the algorithms in our model.

Programming

Come to class on Monday, February 21, 2022 prepared to show me a program that you have written.

Work with the datasets that we saw in Chapters 2 and 3. (California Housing Prices and MNIST, respectively).

Borrow code. Cut and paste. Build on the programs that I have shared with you or on the code that Aurélien Géron has given us. I am looking for evidence that you understand the code. You do not need to show me a lot of original code.

Explain what your program does in comments (or in markdown cells if you are working in a Jupyter notebook.)

Do not try to do too much. Quality matters more than quantity.

You might, for example,...

- Use a class other than `LinearRegression` to predict housing prices.
- Use a class other than `SGDClassifier` for a binary classification of the MNIST digits.
- Experiment with multi-class classification.
- Compute statistics that measure performance. Print or plot these measures.
- Add code for cross-validation.
- Add code that uses a grid search to find good values for hyperparameters.
- You might like to try one of the exercises at the end of Chapter 2 or Chapter 3.

You may choose to write a Python program (a `*.py` file) or a Jupyter notebook (a `*.ipynb` file).