Lecture notes on Udacity's "Introduction to Machine Learning Class"

Lauren E. Blake

May 25, 2017

The class contains 16 lessons as well as a Final Project.

0.1 Lesson 1: Welcome

0.1.1 Introduction I

- Lots of applications for machine learning across diverse fields!

0.1.2 Introduction II

- Keep your eyes out for applications of machine learning and data sets that you could use machine learning on.

0.1.3 Introduction III

- Format: Lectures with quizzes, mini-projects at the end of each lesson.
- Final project at the end that ties together different aspects of the mini-projects.

0.2 Lesson 2: Naive Bayes

0.2.1 ML in the Google Self-Driving Car

- Train the self-driving car by showing car how to drive and giving the computer examples of humans driving. This is an example of a supervised classification problem.

0.2.2 Acerous versus non-Acerous Quiz

- Gave example of acerous animals and non-acerous examples. Then asked whether we thought a horse was acerous or not.
- Answer: A horse is acerous (lacking horns or antlers).

0.2.3 Supervised Classification Examples Quiz

- Need to have training data and then making predictions, recommendations, etc. based on the training set.

0.2.4 Features and Labels Musical Example Quiz

- With a song, features could be tempo, intensity, gender of person singing it, etc. Labels are whether a person likes the song or not.

0.2.5 Features Visualization Quiz

- Answer: She likes those because they are close to the other data points representing songs that the person also likes.

0.2.6 Classification by Eye Quiz

- Answer: Unclear because the new data point is close to two different labels.

0.2.7 Intro To Stanley Terrain Classification

- Features: speed and ruggedness

0.2.8 Speed Scatterplot: Grade and Bumpiness Quiz

- Answer: From the picture, we can see that the terrain looks flat and smooth, particularly relative to the other pictures.

0.2.9 Speed Scatterplot 2

- Answer: From the picture, we can see that the terrain looks very steep and medium bumpy.

0.2.10 Speed Scatterplot 3

- Answer: From the picture, we can see that the terrain looks flat and very.

0.2.11 From Scatterplots to Predictions

- Answer: The points look closer to the blue circles than the red X's.

0.2.12 From Scatterplots to Predictions 2

- Answer: Unclear because the points look equally close to the blue circles than the red X's.

0.2.13 Scatterplots to Decision Surface Quiz

- A decision surface parses out the training data into different features so that a data point falling on one side of the decision surface has one label and on the other side, a different label.
- Answer: Red cross because it is on the same side as the training data with red crosses.

0.2.14 A Good Linear Decision Surface

- When the decision surface is a straight line, it is a "linear" decision surface.
- Answer: Select the line that clearly and consistently separates the red crosses from the blue circles.

0.2.15 Transition into Using Naive Bayes

- Naive Bayes is a common algorithm to find the decision surface.

0.2.16 NB Decision Boundary in Python

- Have 750 training data points and make a decision boundary.

0.2.17 Getting Started with sklearn

- Documentation on Naive Bayes with derivation and code
- clf = GaussianNB() # Create Gaussian classifier

- clf.fit(features, labels) # Fit Gaussian classifier
- clf.predict #Give it a point and get out a label

0.2.18 Gaussian NB Example

- Goes through each line of code in the example section of http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html.

0.2.19 GaussianNB Deployment on Terrain Data

- Answer: Add the following code:

```
#Specify classifier type
clf = GaussianNB()


# Fit the decision boundary
clf.fit(features_train, labels_train)
```

0.2.20 Finding Naive Bayes Accuracy

- accuracy = number of points classified correctly / all points in the test set
- Answer: See Lesson_2_Section_20_Quiz:Calculating_NB_Accuracy on my Udacity Machine Learning GitHub repo.

0.2.21 Training and Testing Data

- Important to train and test on different data sets (need to generalize to new data sets)

0.2.22 Unpacking NB Using Bayes Rule

- What is Naive Bayes?

0.2.23 Cancer Test

- Answer: We can see in from the diagram that the probability that someone with a positive cancer test actually has the disease is approximately 8%.

0.2.24 Normalizing 1

- Answer: The normalizing constant is 0.009+0.099 = 0.108.

0.2.25 Normalizing 2

- Answer: 0.08333. Divide the cancer joint by the normalizing constant to get the posterior.

0.2.26 Normalizing 3

- Answer: 0.916666. Divide the non-cancer joint by the normalizing constant to get the posterior.

0.2.27 Total Probability Quiz

- Answer: 1. Add the answers from Normalizing 2 and Normalizing 3 together.

0.2.28 Bayes Rule Diagram

$$Total probability = 1 = P(\frac{C \mid Pos}{Normalizing constant}) + P(\frac{non - C \mid Pos}{Normalizing constant}) \tag{1}$$

0.2.29 Bayes Rule for Classification

- Who is the person that is sending the email based on the words that they used?

0.2.30 Chris or Sara Quiz

- Answer 1: Sara because the email contains words that she uses with higher probability than Chris.
- Answer 2: Chris because the email contains words that she uses with higher probability than Sara.

0.2.31 Posterior probabilities

- Answers: 0.5714; 1-0.5714; First find

$$Prob\frac{\big(Chris \mid Emailcontainsthewords``lifedeal''\big)}{Constant} \tag{2}$$

Then, find its complement.

0.2.32 Bayesian probabilities on your own

- Prob $\big($Chris | Email contains the words "love deal" $\big)$ =

  Prob $\big($"love deal" | C $\big)$*P$\big($C $\big)$ divided by a normalizing constant

  where the constant is Prob $\big($"love deal" | C $\big)$ P$\big($C$\big)$ + Prob $\big($"love deal" | S $\big)$ P$\big($S$\big)$

  Then, take the complement.

- Answers: 0.5555; 0.4444.

0.2.33 Why Is Naive Bayes Naive

- Don't see underlying process (e.g. who is using the words) but get to see the outcome (e.g. the words that the person used).
- Answer: Word order is being ignored in Bayes Theorem whereas the words used and the length of the message (in terms of which words are used) are used.

0.2.34 Naive Bayes Strengths and Weaknesses

- Pros: Easy to implement, efficient
- Cons: Can break (e.g. phrases with distinct meanings)
- Good for text classification because can treat each word as a feature.

0.2.35 Naive Bayes Mini-Project

- Identify who authored a piece of text (e.g. Chris and Sara) based on previous emails.