Week 3: Bayesian methods

Week 2 quiz review

- NA’s were just missing rows at the end of the file
- Real NA’s were -200 – after dropping fully NA rows, replace -200 with NA
- Replace commas with periods
- Convert times to seconds since epoch (seconds since 1-1-1970
- May end up as the week 1 assignment next time I teach this – ended up being quite difficult, but useful skills

Week 2 Review

- Train/test/validation split – why?
  -
- What is a high-bias model? High-variance?
  -
- Cross-validation
  -
- KNN algorithm
  -

Week 2 Review

- Train/test/validation split – why?
  - We want to compare different models and hyperparameters to pick the best ones, so train on training and test on an unseen “holdout” test set (or validation set in some cases)
- What is a high-bias model? High-variance?
  - High bias means it is too general – e.g., predicting the mean for regression
  - High variance means it is overfitting – fits the training perfectly, but does terribly on the test
- Cross-validation
  - Split data into x parts (say 3), train on 2, test on 1, and try all permutations
- KNN algorithm
  - Training saves all points
  - Evaluation:
    - Calculate distance from test point to all other points
    - Find closest k points, take average for regression, majority vote for classification

R packages

- A major difference between Python and R
  - Python mainly has sklearn for ML
  - R has many different packages; caret is the closest to sklearn in completeness
- Anyone using ‘class’ package for KNN and not caret?

R packages

- Who is the creator of caret and who made `class`?

R packages

- Who is the creator of caret and who made `class`?

Author of class, Brian Ripley:

Author of caret, Max Kuhn

Week 2 review quiz

- KNN classification on Iris dataset
- Turn into Week 2 quiz folder under ‘assignments’ in worldclass
- Participation grade, next week’s quiz will be graded

History of Bayes' Law (Bayes Theorem)

- Reverand Thomas Bayes introduced the idea in 1763
- Predict the future based on our "priors" (prior known probabilities from data)
- Posterier is P(A|B)
- Pronounced 'probability of A given B'

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Bayes' Law (Bayes Theorem)

- Example: customer visiting webpage to sign up for subscription to a service
- $P(A|B)$ = probability of A given B
  - e.g., probability of a customer signing up for a subscription if they spent > 1 minute on the page
- $P(B|A)$ = probability of B given A
  - e.g. probability of customer spending > 1min on a webpage if they signed up for a subscription

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Bayes' Law (Bayes Theorem)

- Example: customer visiting webpage to sign up for subscription to a service
- P(A|B) = probability of A given B
  - Thing we are predicting – might popup a 'limited time offer' if the person tries to exit before 1min
- P(B|A) = probability of B given A
  - e.g. measurable based on data – take conversion rate of customers that spent >1min on page

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

Bayes’ Law (Bayes Theorem)

- Example: customer visiting webpage to sign up for subscription to a service
- P(A) = probability of A
  - Measurable by past data – overall conversion rate
- P(B) = probability of B
  - Measureable by past data – % of people spending > 1min on page

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Bayes' Law (Bayes Theorem)

- Example: customer visiting webpage to sign up for subscription to a service
- $P(A) = 25\%$
  - overall conversion rate
- $P(B) = 50\%$
  - % of people spending > 1min on page
- $P(B|A) = 65\%$
  - % of people that spend >1min on page and sign up

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}$$

Bayes' Law (Bayes Theorem)

- Example: customer visiting webpage to sign up for subscription to a service
- 0.65*0.25/0.5 = 0.325
- A better way would be to represent the "prior" data as a distribution
- If interested in bio stuff like drug testing or for cancer or disease, check out wikipedia's example

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Bayes' Law for classification

- Classic example of machine learning - Naive Bayes classifier for email spam classification
  - We use probability of each word for each class, multiply them to get overall probability for a class

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

http://www.saedsayad.com/naive_bayesian.htm

Example: Naive Bayes on the heart disease dataset

Bayesian Networks

- Directed Acyclic Graph (DAG)
- Probabilities and states of features effect probabilities of other features in Bayesian calculations
- Improvement on Naive Bayes model because this is a bit more complex
- Bayesian Network example (adapted from https://www.r-bloggers.com/bayesian-network-in-r-introduction/)

Exercise: Bayesian Network on heart disease data

- Use bnlearn to create a Bayesian network model on the data.
- Plot the model. Do the connections make sense? Remove any that don’t make sense.
- Find some probabilities of events using cpquery and discuss/interpret them.

Project (Homework): Use NB Classifier for spam/ham dataset

- Hint: take a sample (with sample()) of the data to work through first so it runs quickly, then use the full dataset
- Make a train/test set
- Use cross-validation on the train set to try a few different hyperparameter settings for your Naive Bayes classifier, then report the performance results on the test set with a confusion matrix
- Post the results to the week 3 discussion by next class

Taking it further

- Play around with the tm package a bit more and tweak the termdocumentmatrix settings
- Make wordclouds of most common words in spam/ham
- Naive Bayes on other datasets
- Trying it in Python
- Try out Bayesian models on predicting email opens
- Go through the