Cultural Analytics

ENGL 64.05

Fall 2019

Prof. James E. Dobson

Nov 8, 2019

Use and Misuse of Classification

- Machine learning classification algorithms "learn" which features contribute to knowledge to separate classes.

- The abstract notion of the class is defined by the membership of the training data, provided features, the model, and the parameters used for the model.

- We generally want "better" performance metrics—higher classification accuracy.

- We can use the "slag" generated during the learning process to extract/mine meaningful data.

- This is akin to the discovery of word2vec in the residue of training data.
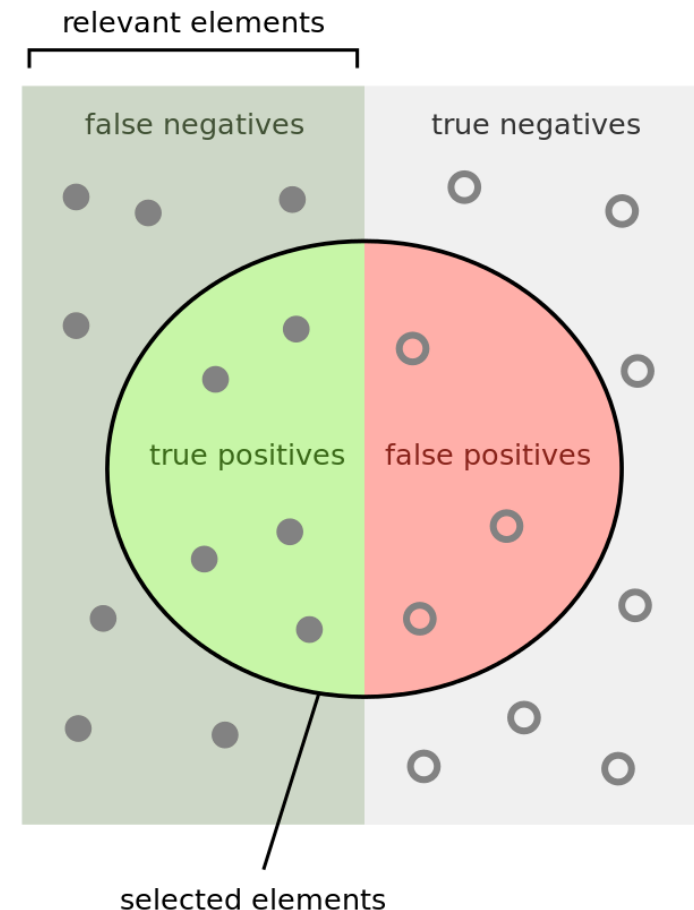
Accuracy / Precision

model accuracy: 0.911

   (91% accurate
   classification
   across model)

```
      precision  recall
bio   0.82       0.89
fic   0.94       0.89
poe   0.97       0.97
```

Precision is the percentage of classifications that are relevant.
Recall is the percentage of total relevant results correctly classified.

relevant elements

false negatives   true negatives

true positives   false positives

selected elements

How many selected items are relevant?

How many relevant items are selected?

Precision =

Recall =

The Confusion Matrix

```
         BIO FIC POE
BIO     [32     3      1]
FIC     [ 6    49      0]
POE     [ 1     0     32]
```

Look for high values in the diagonal
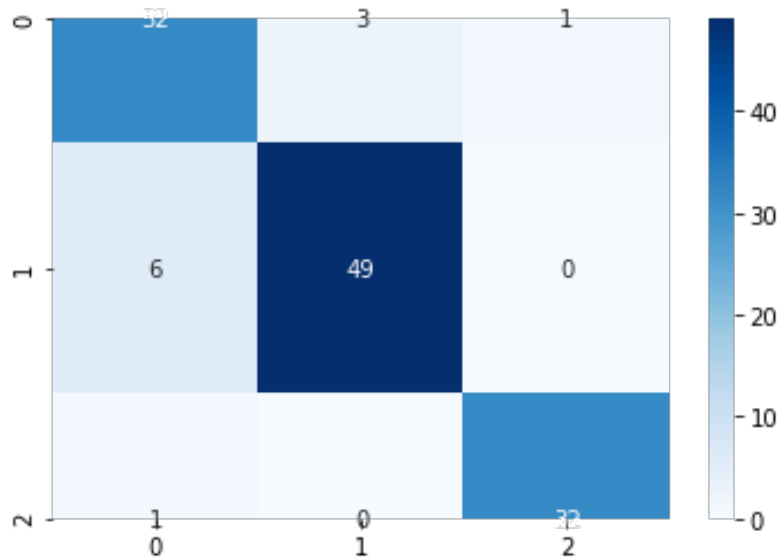
What features contributed the most to our model?

- **bio**: :, vous, for, amnon, joab, cardinal, majesty, book, unto, this, is,), be, saul, in, absalom, |romannumeral|, *, did, to, ;, lord, king, david, of

- **fic**: robin, an, didymus, i, fo, princess, —, that, knights, beatrice, being, it, knight, theodora, george, had, -, don, not, he, him, her, you, powel, a

- **poe**: your, (, thou, hell, some, or, ', their, my, with, then, ■, and, !, —, ,., like, |arabic3digit|, j, all, his, zeus, |'s|, thy, ,

Haber: Hopeful

- Computation can be used to resist normative categories: "It is this self-shattering, or perhaps more specifically a redistribution of the self, that lies at the heart of my vision for a queer method" (162).

- Quantification can make new connections: "All methods, those understood as quantitative and those called qualitative, can be used to categorize and determine, to make and unmake populations, and to determine life chances. But in dismissing quantification we miss new possibilities for both articulating and cultivating queer circuits of connection and sensation" (164).

Rouvroy: Pessimistic

- **Key term**: algorithmic governmentality.
- She's not necessarily opposed to statistics and quantification—these procedures work and are well known.
- She's deeply concerned with models that do not need to concern themselves with people.
- When these models enter into the governing practices of everyday life, we have no way to contest the model, no mode of redress.

Next Class:
The Presentation of Data

- Edward Tufte

- Klein and Posner