CS 165A Discussion 3

--Break Through MP1

Lei Xu

Feb 1

Announcement

- HW2 was out, start early !

Agenda

- Gaussian Naïve Bayes Classifier with Bag-of-Words
- Gaussian Naïve Bayes Classifier with TF-IDF
- Multinomial Naïve Naïve Bayes Classifier with Bag-of-Words

Our task: Movies Review Classification

Compare P (positive | review) and P (negative | review)

Apply the Bayes Rule

$$P(C_i \mid review) = \frac{P(review \mid C_i)\, P(C_i)}{P(review)}$$

Apply the Bayes Rule

$$P(C_i \mid review) = \frac{P(review \mid C_i)\, P(C_i)}{P(review)}$$

How do we know $P(review\ |C_i)$?

1. A Gaussian Perspective

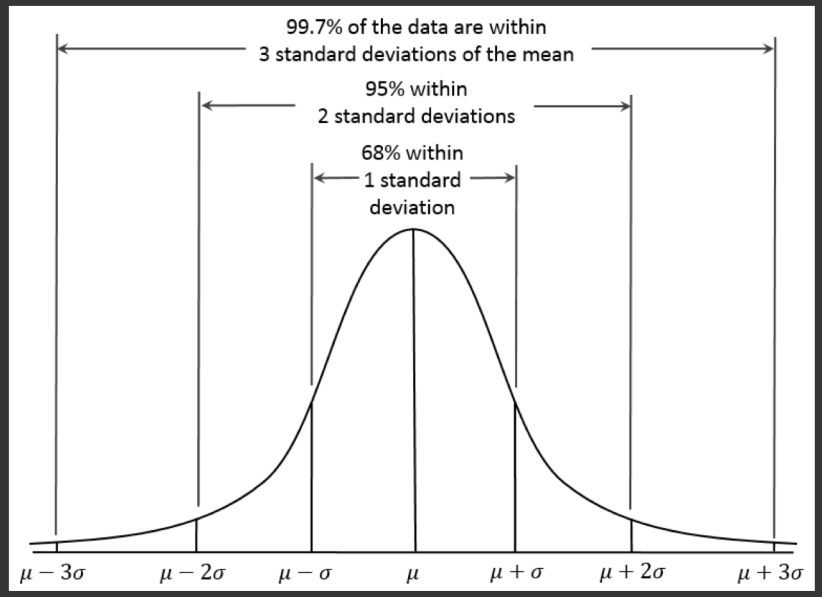
- Assume the probability P(review | C_i) follows a **Multivariate** Gaussian Distribution.

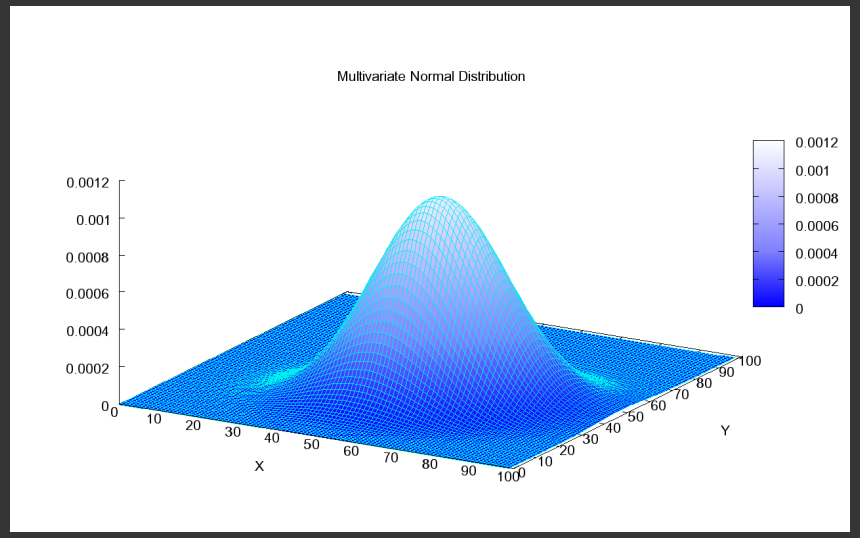
- Use Feature Vector represent each review

$\boldsymbol{X}\ (x_1, x_2\ \ldots x_n\ )$ ~ N (μ, Σ)

μ: Mean

Σ: Covariance Matrix





The Multivariate Gaussian Distribution

$$p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$$

How do we know μ, Σ ?

Maximum Likelihood Estimate a.k.a. learn it from the training data.

Find the best parameter μ, Σ that fit the data.

How to calculate this ?

Try it out yourself in HW 2 :P

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \hat{\mu}) (x_n - \hat{\mu})^T .$$

Feature Vector *x*

- Bag-of-Words

Eg. In positive class
Review 1 :" another movie about trip "
Review 2: "a very very great movie "

	a	great	another	movie	trip	about	very
review 1	0	0	1	1	1	1	0
review 2	1	1	0	1	0	0	2

x1 = (0,0,1,1,1,1,0)
X2 = (1,1,0,1,0,0,2)

• TF-IDF Feature

- Recall in the BoW feature, each review is represented by a vector.
- Vector contains the frequency of each word in the vocabulary.
- Replace this frequency counts with TF-IDF score for each word.
- TF = (Number of times $word_i$ appears in the review) / (Total number of words in this review)
- IDF = log(Total number of reviews / Number of reviews with $word_i$ in it)

TF * IDF

Multinomial Naïve Bayes Classifier

see lecture slide6 - 36

- $P(review\ |Positive) = \prod P(word_i|Positive)$

$$P(word_i \mid Positive) = \frac{\#\ of word_i\ in\ Positive}{total\ \#\ of\ words\ in\ entire\ Positive\ data}$$

- Multiply by zero --- Apply Laplace Smoothing
- Underflow --- Take log term

Summary

- Use training data, build your vocabulary for two classes, separately.
- Estimate the parameter of your Gaussian distribution
  - Use BoW as feature
  - Use TF-IDF as feature
- Or use Multinomial Naïve Bayes model
- Classify the testing data according to Naïve Bayes Rule.

Thanks !

Happy Chinese New Year ! （Feb 5)