Algorithms for Machine Learning

Second Week
Chiranjib Bhattacharyya

Dept of CSA, IISc
chibha@chalmers.se

January 25, 2012

Agenda

- Bayes Classifier (continued)
- Naive Bayes Classifier

Probability Distributions

Bernoulli

$$P(X=1)=p,\ P(X=0)=1-p\ X\sim Ber(p)$$

Gaussian Distribution

$$P(X=x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ X\sim N(\mu,\sigma^2)$$

Multivariate Gaussian

$$P(X=x)=\frac{1}{\sqrt{(2\pi)^d|\Sigma|}}e^{-\frac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)}\ X\sim N(\mu,\Sigma)$$

Maximum Likelihood

Let $Z \sim P(\theta)$ Estimate θ from n i.i.d observations of Z

Maximum Likelihood

Let $Z \sim P(\theta)$ Estimate θ from n i.i.d observations of Z

The maximum likelihood estimate

$$\mathscr{L}\mathscr{L}(\theta) = \sum_{i=1}^{n} logP(Z_i = z_i|\theta)$$

$$\theta^* = \text{argmax}_\theta \mathscr{L}\mathscr{L}(\theta)$$

Maximum Likelihood

Bernoulli

Given $x_1, x_2, \ldots, x_n \sim Ber(p)$ $\hat{p} = \frac{1}{n}\#(x_i = 1)$

Gaussian Distribution

Given $x_1, x_2, \ldots, x_n \sim N(\mu, \sigma^2)$ then

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \;\; \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

Multivariate Gaussian Distribution

Given $x_1, x_2, \ldots, x_n \sim N(\mu, \Sigma)$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \;\; \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})^\top$$

Bayes Classifier

From Data estimate $P(X = x|Y = i)$ and $P(Y = i)$.
Compute $P(Y = i|X = x) = \frac{P(X=x|Y=i)P(Y=i)}{P(X=x)}$

$$score_i(x) = logP(Y = i|X = x)$$

$$= logP(X = x|Y = i) + c_i$$

$$c_i = logP(Y = i) - logP(X = x)$$

Bayes Classifier

$$f(x) = argmax_i score_i(x)$$

Independent features, gaussian distributed

Assumptions

$x = (f_1, f_2, \ldots, f_d)^\top$

$$x \in \mathbb{R}^d \; P(X = x | Y = 1) = \prod_{j=1}^{d} N(\mu_{1j}, \sigma_{1j}^2)$$

Independent features, gaussian distributed

Assumptions

$x = (f_1, f_2, \ldots, f_d)^\top$

$$x \in \mathbb{R}^d \; P(X = x | Y = 1) = \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\sigma_{1j}^2}} e^{-\frac{1}{2\sigma_{1j}^2}(f_j - \mu_{1j})^2}$$

Independent features, gaussian distributed

Assumptions

$x = (f_1, f_2, \ldots, f_d)^\top$

$$x \in \mathbb{R}^d \; P(X = x | Y = 1) = \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\sigma_{1j}^2}} e^{-\frac{1}{2\sigma_{1j}^2}(f_j - \mu_{1j})^2}$$

$$score_1(x) = -\frac{1}{2} \sum_{j=1}^{d} \left(log(2\pi\sigma_{1j}^2) + \frac{(f_j - \mu_{1j})^2}{\sigma_{1j}^2} \right) + c_1$$

Similarly for class 2.

Naive Bayes Classifier

- Assumption on feature independence
- works well for many problems, specially on *text classification*

Spam Emails

Search Results | Delete

Subject: Your Email Address Has Made You A Millionaire!!!
From: "The Awards Committee" <candexinfo@bellnet.ca>
Date: Tue, September 9, 2008 8:11 am
To: winners@euromillions.org
Priority: Normal
Options: View Full Header | View Printable Version | Download this as a file

You Have Been Selected As Winner Of A Cash Prize

Of One Million Euro (€1,000,000.00 EURO). For More

Information Contact Mr. Donald Wong, Email:donaldwong.anzbnk@live.com

Regards,

Ms. Roxanne Presley.

Spam Emails

Current Folder: **INBOX**

Compose Addresses Folders Options Search Help

Search Results | Delete Forv

Subject: AWARD NOTIFICATION
From: "UK NATIONAL LOTTERY" <info@winners.com>
Date: Wed, March 25, 2009 6:32 pm
Priority: Normal
Options: View Full Header | View Printable Version | Download this as a file

```
P.O.Box 1010
Liverpool
L70 1NL
United Kingdom.
Ref: XYL /26510460037/05
Batch: 24/00319/IPD
Ticket Number : 56475600545188                          AWARD NOTIFICATION

This is to inform you that you have been  selected for a cash prize
of£1,500,000.00 pounds held  on  the 25th MARCH,2009 in London UK.
The selection process was carried out through random selection
Ourcomputerized email selection system(ess)
from a database of over 250,000  email Addresses drawn from which you were
selected.

You are to contact the fiduciary claims department by your personal
information with e-mail Giving Below;
```

Naive Bayes Classifier: Bernoulli model

Create a feature list where each feature is on/off. Denote the feature map $x = [f_1, \ldots, f_d]^\top$
$P(X = x|Y = y) = \prod_{i=1}^{d} P(F_i = f_i|Y = y)$

$$p_{1i} = P(F_i = 1|Y = 1) \; p_{2i} = P(F_i = 1|Y = 2)$$

$$score_1(x) = \sum_i (f_i log p_{1i} + (1 - f_i) log(1 - p_{1i})) + c_1$$

similarly $score_2(x)$
Bayes Classifier: Output the class with the higher score

Naive Bayes: Bernoulli

Source: Introduction to Information Retrieval. (Manning, Raghavan, Schutze)

```
TRAINBERNOULLINB(ℂ, 𝔻)
1  V ← EXTRACTVOCABULARY(𝔻)
2  N ← COUNTDOCS(𝔻)
3  for each c ∈ ℂ
4  do N_c ← COUNTDOCSINCLASS(𝔻, c)
5      prior[c] ← N_c / N
6      for each t ∈ V
7      do N_ct ← COUNTDOCSINCLASSCONTAININGTERM(𝔻, c, t)
8          condprob[t][c] ← (N_ct + 1)/(N_c + 2)
9  return V, prior, condprob

APPLYBERNOULLINB(ℂ, V, prior, condprob, d)
1  V_d ← EXTRACTTERMSFROMDOC(V, d)
2  for each c ∈ ℂ
3  do score[c] ← log prior[c]
4      for each t ∈ V
5      do if t ∈ V_d
6            then score[c] += log condprob[t][c]
7            else score[c] += log(1 − condprob[t][c])
8  return arg max_{c∈ℂ} score[c]
```

▶ **Figure 13.1** NB algorithm (Bernoulli model): Training and testing. The add-one smoothing in Line 8 (top) is in analogy to Equation 119 with $B = 2$.

Discriminant functions

Bayes Classifier

$$h(x) = \text{sign}\left(\sum_{i=1}^{d} f_i w_i - b\right)$$

$w_i = \log \frac{p_{1i}(1-p_{2i})}{(1-p_{1i})p_{2i}}$

$h(x)$ is sometimes called Discriminant function

Gaussian class conditional distributions

Let the class conditional distributions be $N(\mu_1, \Sigma)$ and $N(\mu_2, \Sigma)$.
The Bayes classifier is given by

$$h(x) = \mathrm{sign}(w^\top x - b)$$

$w = \Sigma^{-1}(\mu_1 - \mu_2)$

Linear classifiers

L

inear Classifiers can be Bayes classifier

$$h(x) = \text{sign}(w^\top x - b)$$

Naive Bayes: Bernoulli
Gaussian class conditional distribution with same covariance

Bayes Classifier

- Gives the best *Generalization* error
- Computing $P(X = x | Y = y)$ is hard
- Easy under severe assumptions on the distribution