Announcements

Assignments:

- HW4
  - Release delayed to tomorrow
  - Due date delayed to Thu, 2/20, 11:59 pm

Midterm Conflicts

- See Piazza post

Plan

Last time

- Wrap up MLE vs MAP
- Intro to Naïve Bayes

Today

- MLE vs MAP
- Naïve Bayes Assumptions
- Naïve Bayes MLE
- Naïve Bayes MAP
- Generative Models

Introduction to Machine Learning

Generative Models

Instructor: Pat Virtue

SPAM Detection Handout

Previous Piazza Poll

What method were we using to estimate parameters in our Naïve Bayes handout?

Generative vs Discriminative

MLE vs MAP vs Generative vs Discriminative

SPAM Detection Data and Assumptions

Naïve Bayes MLE

Whiteboard

Naïve Bayes MLE

$$\mathcal{D} = \{y^{(n)}, \boldsymbol{x}^{(n)}\}_{n=1}^{N}$$
$$y^{(n)} \in \{0,1\}$$
$$\boldsymbol{x}^{(n)} \in \{0,1\}^{M}$$
$$\phi \in [0,1]$$
$$\boldsymbol{\Theta} \in [0,1]^{Mx2}$$

$$L(\phi, \boldsymbol{\Theta}) = p(\mathcal{D} \mid \phi, \boldsymbol{\Theta})$$

$$= \prod_{n=1}^{N} p\left(\mathcal{D}^{(n)} \mid \phi, \boldsymbol{\Theta}\right)$$ i.i.d assumption

$$= \prod_{n=1}^{N} p\left(y^{(n)}, \boldsymbol{x}^{(n)} \mid \phi, \boldsymbol{\Theta}\right)$$

$$= \prod_{n=1}^{N} p\left(y^{(n)} \mid \phi\right) p\left(\boldsymbol{x}^{(n)} \mid y^{(n)}, \boldsymbol{\Theta}\right)$$ Generative model

$$= \prod_{n=1}^{N} p\left(y^{(n)} \mid \phi\right) p\left(x_1^{(n)}, x_2^{(n)}, \ldots, x_M^{(n)} \mid y^{(n)}, \boldsymbol{\Theta}\right)$$

$$= \prod_{n=1}^{N} p\left(y^{(n)} \mid \phi\right) \prod_{m=1}^{M} p\left(x_m^{(n)} \mid y^{(n)}, \theta_{m,y}\right)$$ Naïve Bayes

$$= \prod_{n=1}^{N} \phi^{y^{(n)}} (1-\phi)^{1-y^{(n)}} \prod_{m=1}^{M} \theta_{m,1}^{\mathbb{I}\left(y^{(n)}=1 \,\wedge\, x_m^{(n)}=1\right)} \left(1-\theta_{m,1}\right)^{\mathbb{I}\left(y^{(n)}=1 \,\wedge\, x_m^{(n)}=0\right)}$$
$$\theta_{m,0}^{\mathbb{I}\left(y^{(n)}=0 \,\wedge\, x_m^{(n)}=1\right)} \left(1-\theta_{m,0}\right)^{\mathbb{I}\left(y^{(n)}=0 \,\wedge\, x_m^{(n)}=0\right)}$$

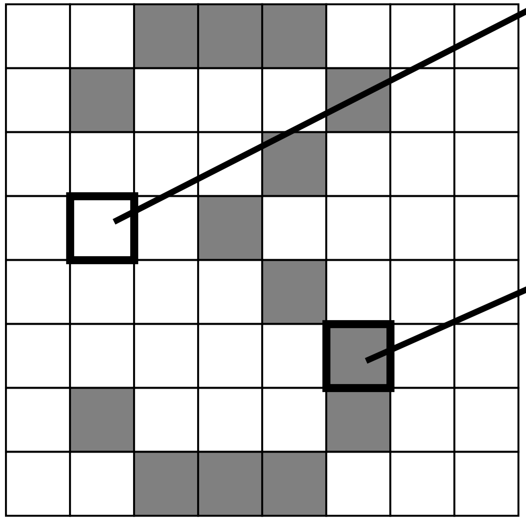
$$= \phi^{N_{y=1}} (1-\phi)^{N_{y=0}} \prod_{m=1}^{M} \theta_{m,1}^{N_{y=1,x_m=1}} \left(1-\theta_{m,1}\right)^{N_{y=1,x_m=0}} \theta_{m,0}^{N_{y=0,x_m=1}} \left(1-\theta_{m,0}\right)^{N_{y=0,x_m=0}}$$

Naïve Bayes MAP

Laplace Smoothing

Naïve Bayes for Digits

y	$p(Y)$
1	0.1
2	0.1
3	0.1
4	0.1
5	0.1
6	0.1
7	0.1
8	0.1
9	0.1
0	0.1



y	$p(X_{3,1} = 1 \mid y)$
1	0.01
2	0.05
3	0.05
4	0.30
5	0.80
6	0.90
7	0.05
8	0.60
9	0.50
0	0.80

y	$p(X_{5,5} = 1 \mid y)$
1	0.05
2	0.01
3	0.90
4	0.80
5	0.90
6	0.90
7	0.25
8	0.85
9	0.60
0	0.80

Generative Models with Continuous Features

Bernoulli class distribution with Gaussian class-conditional distribution