CS 360: Machine Learning

Sara Mathieson, Sorelle Friedler

Spring 2024

HAVERFORD COLLEGE

Sit somewhere new!

Admin

- **EVERYONE:** Sign in again

- We will be making waitlist decisions and lab switches after looking at attendance today

- Sorelle office hours **TODAY: 3-4pm in H110**

- Sara office hours **Monday: 3-4pm in Zubrow**

- TA hour schedule hopefully up today on Piazza

Candidate lunch/talk today!

- **Talk**: “Routing our Way to Next-Generation Internet Services”
  - 4:15 p.m. with a tea reception at 4:00 p.m.
  - H109
- **Student lunch**
  - 12:30-1:30pm
  - Swarthmore Room of the DC

Class notes and materials

- Sorelle is taking **notes** and we will post these along with **slides, board photos, and handouts**

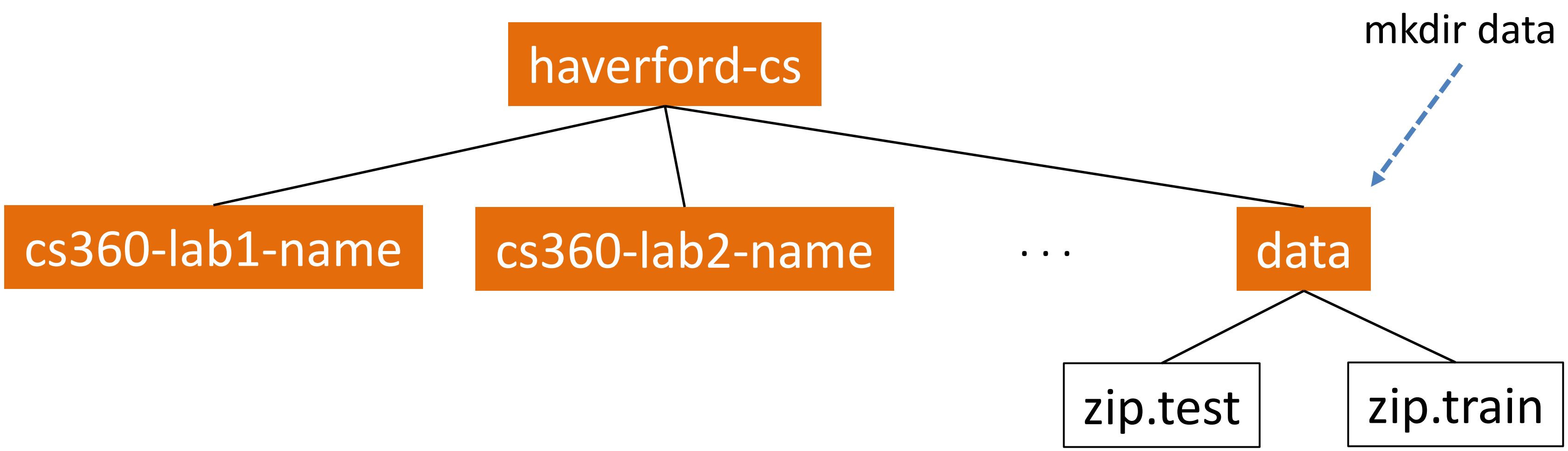
- Opportunity for extra credit (rare!) for creating **video solutions for handouts**
  - I will post details on Piazza and you can sign up for one handout
  - I will attempt to create video and/or written solutions for the rest

Reading for Week 1

- Geron Chap 3 through pg 119 (binary classification and evaluation metrics)
- Geron Chap 4 through pg 151 (linear regression and gradient descent)
- Geron Chap 4 pg 164-169 (logistic regression)

Unix/GitHub workflow issues

- Come to Sorelle's office hours today if:
  - You can't clone or push
  - You cannot get the zipcode data



python3 lab01.py -r ../data/zip.train -e ../data/zip.train -d 5

Unix/GitHub workflow issues

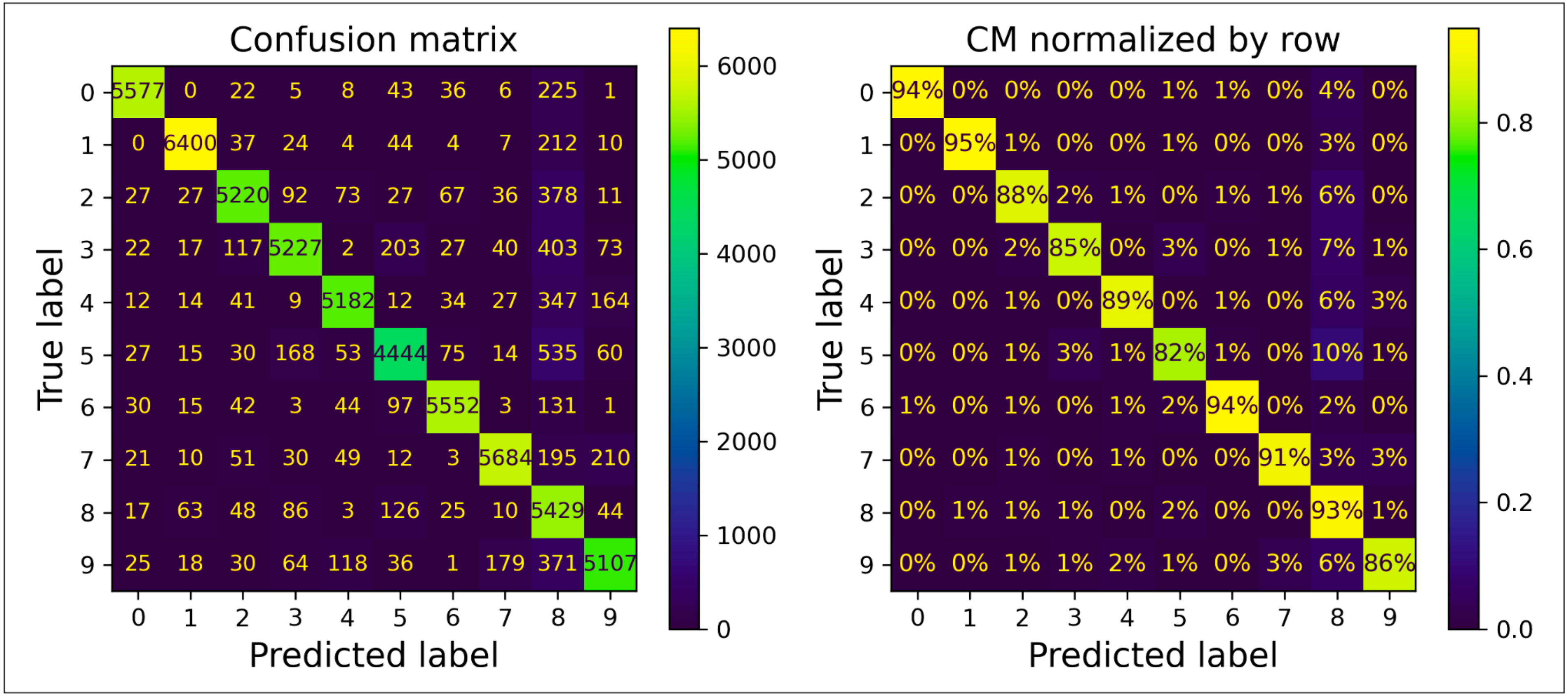
scp <source> <dest>

scp

~~smathieson~~@cook.cs.haverford.edu:/home/smathieson/Public/cs360/zip/zip.train

~~~/Desktop/~~zip.train
~~~

Another example of confusion matrix <u>normalization</u>



Geron Fig 3-9

Outline for Jan 25

- Naïve Bayes review
- ROC curve review
- Logistic Regression review
- Gradient descent review

Outline for Jan 25

- Naïve Bayes review
- ROC curve review
- Logistic Regression review
- Gradient descent review

Informal quiz: discuss with a partner

- Naïve Bayes question from Handout 1

(a) If $\vec{x}$ is a vector of *features* and y is the associated *label*, the components of a typical Bayesian model are $p(\vec{x})$, $p(y)$, $p(y|\vec{x})$, $p(\vec{x}|y)$. What are the terms for each of these probabilities?

p(x)	evidence
p(y)	prior
p(y|x)	posterior
p(x|y)	likelihood

(b) Using Bayes' rule, arrange these probabilities into an equation.

p(x) p(y|x) = p(y) p(x|y) => valid form of Bayes' rule!

p(y|x) = p(y) p(x|y) / p(x) => more common

(c) If $\vec{x} = [x_1, x_2]^T$ (two features or $p = 2$), how would you approximate $p(\vec{x}|y)$ using our Naive Bayes assumption?

$p(x|y) = p(x_1|y)\, p(x_2|y)$

$$p(A,B) = p(A)\,p(B|A)$$

Naive Bayes

$$k \in \{1, 2, 3 \ldots K\}$$

$$p(y=k|\vec{x}) = \frac{p(y=k)\,p(\vec{x}|y=k)}{p(\vec{x})}$$

prediction

$$\hat{y} = \underset{k=1 \ldots K}{\operatorname{argmax}} \; \frac{p(y=k)\,p(\vec{x}|y=k)}{\cancel{p(\vec{x})}}$$

likelihood

$$p(\vec{x}|y=k) = p(x_1, x_2, \ldots, x_p | y$$

$$\approx p(x_1|y)\,p(x_2|y$$

Naive Bayes assumption

$\Rightarrow$ features are conditionally independent given the label

$$p(A,B) \neq p(A)\,p(B)$$

if independent $\Rightarrow$ then equal

$x_p \mid y=k)$

$x_2 \mid y) \cdots p(x_p \mid y)$

x_1 = four legs

x_2 = fur

y = cat

$$\boxed{p(y=k \mid \vec{x}) \approx p(y=k) \prod_{j=1}^{p} p(x_j \mid y=k)}$$

$p(\vec{x})$

single example

product

$$X = \begin{bmatrix} \vec{x}^T \\ \\ \end{bmatrix}_{n \times p}$$

$\sum$ (sum)

Prior

$$p(y=k) \approx \cancel{\frac{N_k}{n}}$$

N_k ⇒ num train with label k

n ⇒ total num train

$$\boxed{p(y=k) \approx \frac{N_k + 1}{n + K}}$$

$$\Rightarrow \sum_k p(y=k) = 1$$

$$\sum_k \frac{N_k + 1}{n + K} = \frac{n + K}{n + K}$$

k

$\boxed{\text{likelihood}}$

$$p(x_j = v \mid y = k) = \frac{p(x_j = v, y = k)}{p(y = k)}$$

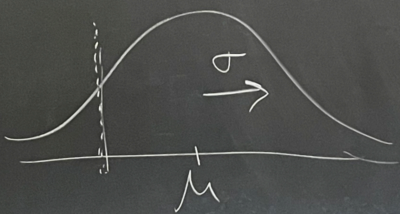
$$\approx \frac{N_{k,j,v} + 1}{N_k + \underbrace{|f_j|}_{\text{\# of possible values for feature } j}}$$

$\boxed{j = 5}$

$\Rightarrow x_5$ or 5th feature

$\Rightarrow x_5 \in \{\text{red}, \text{green}, \text{blue}\}$

$\boxed{v = \text{blue}}$



In practice (binary)

$p = 3$ features

$$\vec{x}_{test} = [x_1 = v_1, x_2, x_3]^T$$

$\boxed{y = 0}$

$$\overbrace{\left(\frac{N_0 + 1}{n + 2}\right)}^{\text{prior}} \overbrace{\left(\frac{N_{0,1,v_1} + 1}{N_0 + 2}\right)}^{\text{feature 1}} \overbrace{\left(\frac{N_{0,2,v_2} + 1}{N_0 + 2}\right)}^{\text{feature 2}} \overbrace{\left(\ldots\right)}^{3}$$

$\boxed{y = 1}$

......

$$p(y=0\mid\vec{x}) \approx 0.015 \Rightarrow \frac{0.015}{0.015+0.0718}$$

$$p(y=1\mid\vec{x}) \approx 0.0718$$

[redacted]

$\Rightarrow$

$$p(y=0\mid\vec{x}) \approx 0.16$$

$$p(y=1\mid\vec{x}) \approx 0.84 \quad \star$$

$t = \cancel{0.3}$ 0.2 $\qquad t = 0.5 \qquad t = \cancel{0.7}$ 0.8

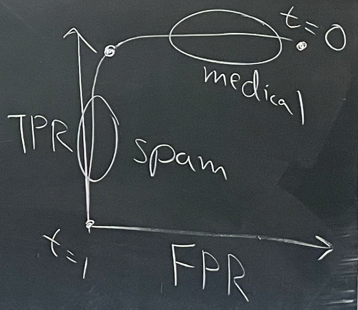
$$p(y=1\mid\vec{x}) \geq t \Rightarrow \text{predict } 1$$

$$\text{o.w.} \Rightarrow \text{predict } 0$$

$\boxed{t = 0.5 \Rightarrow \hat{y} = 1}$

$\boxed{t = 0.9 \Rightarrow \hat{y} = 0}$

$\boxed{t = 0.1 \Rightarrow \hat{y} = 1}$



Handout 2

- Change $t=0.3$ to $t=0.2$
- Change $t=0.7$ to $t=0.8$

①

	$t=0.2$	$t=0.5$	$t=0.8$	y
$y=1$	0	0	0	0
	1	1	1	1
	0	0	0	0
	0	0	0	1
	1	1	1	1
	1	0	0	0
	1	1	0	1
	1	1	1	1
0.2	1	0	0	0
	1	0	0	0

② $t=0.2$ $\quad acc = \frac{6}{10}$

$\boxed{t=0.5 \quad acc = \frac{9}{10}}$

$t=0.8$ $\quad acc = \frac{8}{10}$

④ total $= n = 50$

$$P(y=0) \approx \frac{35+1}{50+2}$$

$$P(y=1) \approx \frac{15+1}{50+2}$$

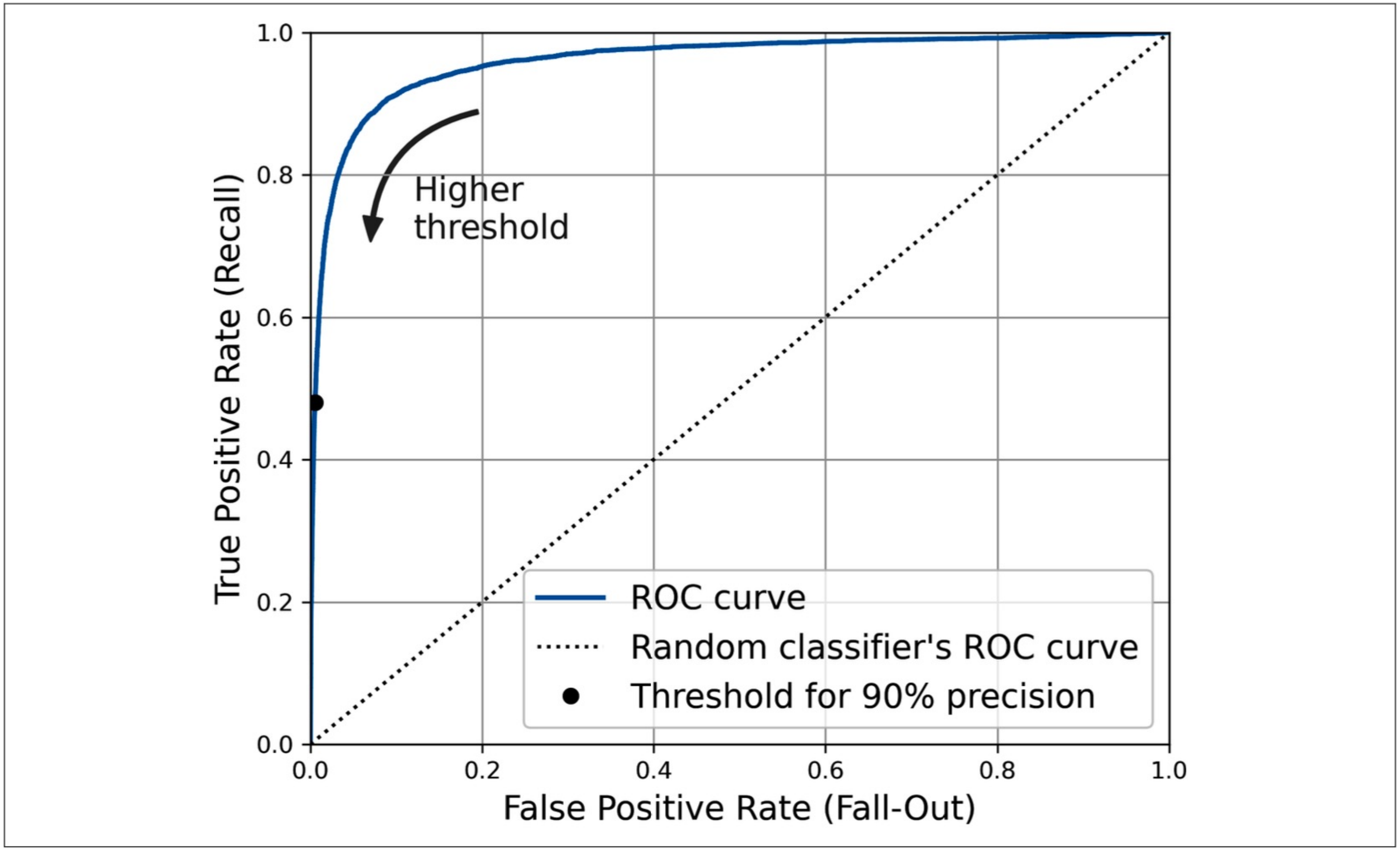
$$p(x_1 = \text{red} \mid y = 0) = \frac{N_{0,1,\text{red}} + 1}{N_0 + 3}$$

$$= \frac{19+1}{35+3}$$

Outline for Jan 25

- Naïve Bayes review
- ROC curve review
- Logistic Regression review
- Gradient descent review

ROC curve example



Geron Fig 3-7