CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019

SWARTHMORE

Admin

- **Lab 3 due Thursday**
  - If you are finished, you do not need to come to lab today (but you are welcome to start on Lab 4)
  - If you are “finished” but don’t understand something, please come to lab and ask!
- **Midterm 1 Feb 27** (in lab)
  - Pick up a study guide!
  - You may use a 1-page (front and back) “cheat-sheet”, but no other resources
- **Lab 4 due March 8** (Friday before spring break)
  - Will run partner script right after class
  - Let me know ASAP of partner preferences!

Outline for February 20

- Naïve Bayes
- Bayes Theorem
- Confusion Matrices

Bayes Rule

oint
"and"

$$\underbrace{P(A,B)}_{\text{rain, no umbrella}} = P(A)P(B|A)$$

$$= P(B)P(A|B)$$

$$\Rightarrow P(A)P(B|A) = P(B)P(A|B)$$

$$\Rightarrow \boxed{P(B|A) = \frac{P(B)P(A|B)}{P(A)}}$$

$$P(\vec{x}\,|\,y=k) = P(\overset{A}{x_1}, \overbrace{x_2, \dots, x_p}^{B} \,|\, y=k)$$

$$= P(x_1 | x_2, \dots, x_p, y=k) \cdot P(x_2, x_3, \dots, x_p \,|\, y=k)$$

$$= P(x_1 | \cancel{x_2 \dots x_p}, y=k)\, P(\underset{A}{x_2} \,|\, \underbrace{\cancel{x_3, \dots}, x_p}_{B}, y=k)\, P(x_3 \dots$$

$$\approx$$

Naive Bayes Assumption

→ feature j is independent of all other features give class label

← Board to the left was unfortunately erased!

$p(x_2, x_3, \ldots, x_p \mid y=k)$

$p(x_2 \mid \cancel{x_3, \ldots, x_p}, y=k)\, p(\underbrace{x_3, \ldots, x_p}_{B} \mid y=k)$

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

...pendent

...atures,

...bel

<u>model</u>

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

"proportional to"

$$\boxed{\hat{y} = \underset{k \in \{1,2,\ldots,K\}}{\operatorname{argmax}}\; p(y=k) \prod_{j=1}^{p} p(x_j \mid y=k)}$$

<u>prediction!</u>

$\underbrace{\theta_k \quad \theta_{j,v,k}}_{\text{plug in}}$

$\overbrace{\quad f_1 \quad f_2 \quad}^{p=2}$

$\vec{X}$	f_1	f_2	Y
$\vec{X}_1$	pos	neg	1
$\vec{X}_2$	pos	pos	2
$\vec{X}_3$	pos	neg	2
$\vec{X}_4$	neg	neg	1
$\vec{X}_5$	pos	neg	2
$\vec{X}_6$	neg	neg	1
$\vec{X}_7$	neg	pos	2

$K=2$

estimate $p(y=k)$

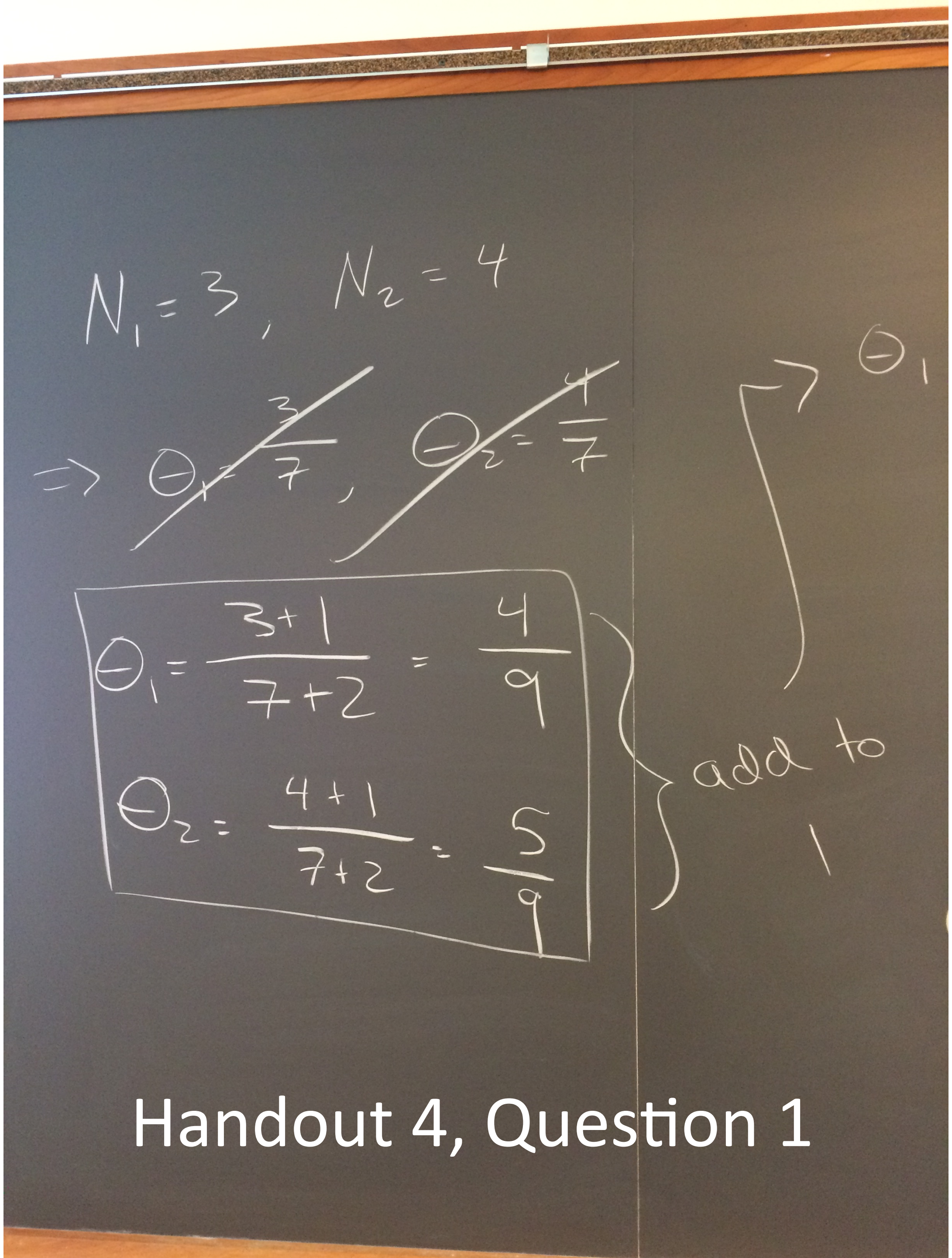
$$\theta_k = \frac{N_k + 1}{n + K}$$

LaPlace counts

N_k = # of datapoints with class k

$N_1 = 3, \quad N_2 = 4$

Handout 4 example



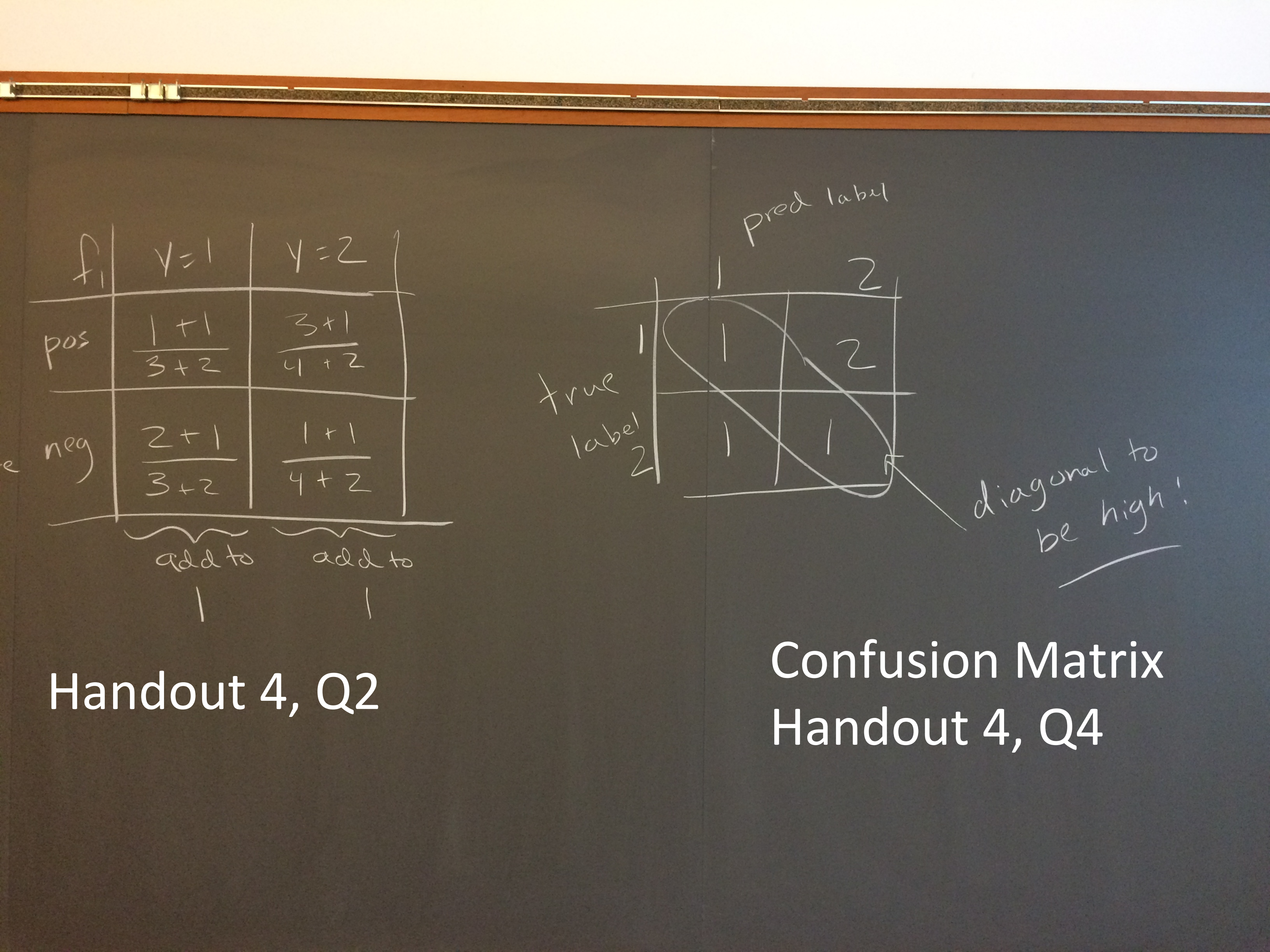
Handout 4, Question 1

$\rightarrow \theta_1 + \theta_2 + \cdots \theta_K = \frac{(N_1 + 1) + (N_2 + 1) \cdots (N_K + 1)}{n + \;?} = \frac{\sum_k N_k + K}{n + K}$

$$\underbrace{\theta_{j,v,k}}_{\text{estimate for } \boxed{P(x_j = v \mid y = k)}} = \frac{N_{j,v,k} + 1}{N_k + |f_j|}$$

\# of possible values feature j can take on

...d to 1



Handout 4, Q2

Confusion Matrix
Handout 4, Q4