

CS 260: Foundations of Data Science

Prof. Sara Mathieson

Spring 2025



HAVERFORD
COLLEGE

- Review today (in class and lab)
- **Exam Thursday**
- Reminder: can use a one-page resource sheet (created by you), but no other resources (i.e. no calculator)

Outline

- Midterm 2 Review
 - Logistic regression and cross entropy
 - Naive Bayes
 - Disparate impact
- Less focus
 - tSNE
 - t-tests

Outline

- Midterm 2 Review
 - Logistic regression and cross entropy
 - Naive Bayes
 - Disparate impact
- Less focus
 - tSNE
 - t-tests

From the study guide

5. Logistic Regression

- Motivation for **logistic regression**; our model is a **logistic function** that takes in $\vec{w} \cdot \vec{x}$
- Logistic regression creates a *linear* decision boundary (compute/visualize for $p = 1$)
- In logistic regression our cost is the **negative log likelihood** (don't need to derive)
- Intuition/visualization of the cost function (and relationship to **cross entropy**)
- **Stochastic gradient descent** (SGD) for logistic regression, relationship to linear regression
- Interpretation of the weights as feature importance

logistic regression cost function

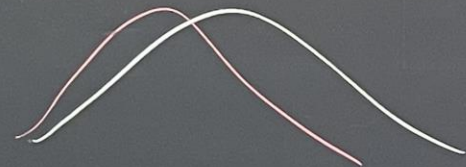
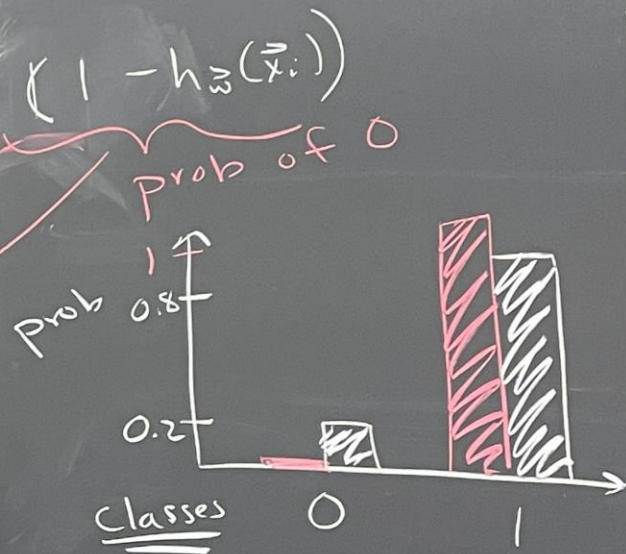
$$J(\vec{w}) = - \sum_{i=1}^n y_i \log \underbrace{h_{\vec{w}}(\vec{x}_i)}_{\text{prob of 1}} + (1-y_i) \log \underbrace{(1-h_{\vec{w}}(\vec{x}_i))}_{\text{prob of 0}}$$

true
 $y_i = 1 \Rightarrow \underbrace{[0, 1]}_{\text{true prob dist}}$

$y_i = 0 \Rightarrow \underbrace{[1, 0]}$

pred prob dist
 $[0.2, 0.8]$
 $1-h_{\vec{w}}(\vec{x}) \quad h_{\vec{w}}(\vec{x})$
 $[0.7, 0.3]$

$$h_{\vec{w}}(\vec{x}) = \frac{1}{1 + e^{-\vec{w} \cdot \vec{x}}} = \underbrace{0.8}_{\text{for example}}$$



$$\vec{w} = \begin{bmatrix} -2 \\ 7 \end{bmatrix} \begin{matrix} \leftarrow w_0 \\ \leftarrow w_1 \end{matrix}$$

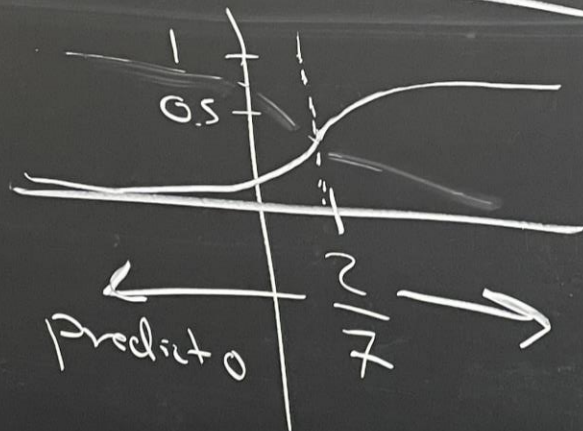
$$-2 + 7x \geq 0 \Rightarrow \text{predict } 1$$

Solve
for x

$$7x \geq 2$$

$$x \geq \frac{2}{7}$$

decision
boundary



train

$$X \rightarrow (n, p)$$

for t in range(T):

for i in range(n):

$$\vec{w} \leftarrow \vec{w} - \eta (h_{\vec{w}}(\vec{x}_i) - y_i) \vec{x}_i$$

check cost

return \vec{w}^*

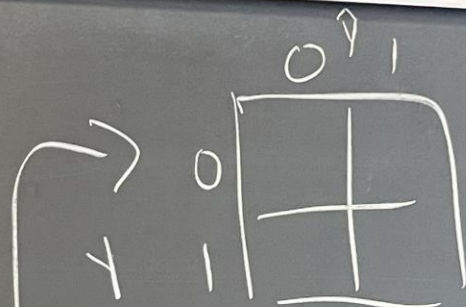
test

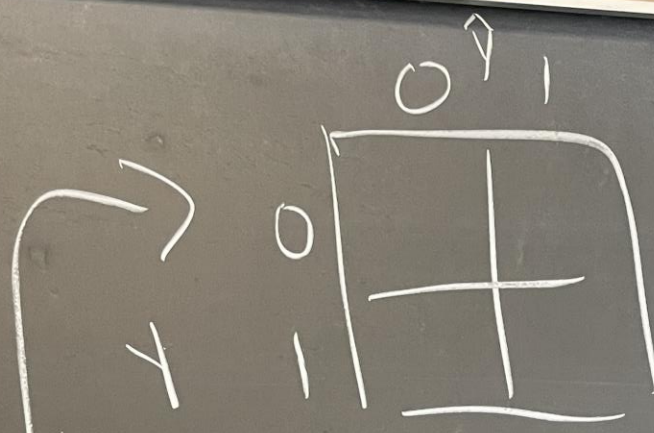
$$X_{\text{test}} \rightarrow (m, p)$$

for \vec{x}_i in X_{test} :

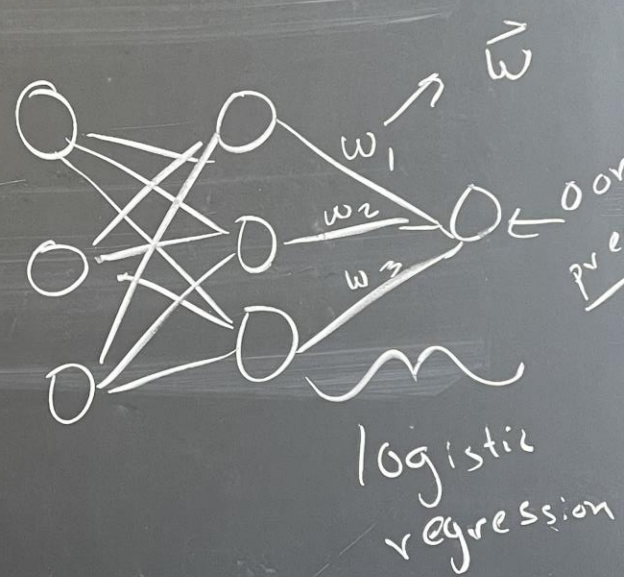
$$\hat{y}_i = 1 \text{ if } h_{\vec{w}}(\vec{x}_i) \geq 0.5 \text{ else } 0$$

shuffle





accuracy,
etc



changes little
changes a lot

Outline

- Midterm 2 Review
 - Logistic regression and cross entropy
 - Naive Bayes
 - Disparate impact
- Less focus
 - tSNE
 - t-tests

From the study guide

2. Naive Bayes

- Bayes rule in data science: identify and explain the [evidence](#), [prior](#), [posterior](#), [likelihood](#).
- Derivation of the [Naive Bayes model](#) for $p(y = k|\vec{x})$ (via the Naive Bayes assumption).
- How do we estimate the probabilities of a Naive Bayes model?
- [Laplace counts](#) (motivation, application details)
- How can we predict the label of a new example after fitting a Naive Bayes model?
- What types of features/label do we currently require for Naive Bayes?
- How Naive Bayes can be implemented using [dictionaries](#) in Python

Naive Bayes

Bayes: $P(A, B) = P(\textcircled{B}) P(\textcircled{A|B})$ } always true

independence: $P(A, B) = P(A) P(B)$ } sometimes true

conditional independence

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

likelihood

$$P(\underbrace{x_1, x_2, x_3}_{3 \text{ features}} | y)$$

$$= P(\textcircled{x_1} | y) P(x_2, x_3 | \textcircled{x_1}, y)$$

↓ assumption

$$= P(x_1 | y) P(\underbrace{x_2, x_3}_{\rightarrow} | y)$$

Naïve Bayes Model

$$p(y = k | \mathbf{x}) \propto p(y = k) \prod_{j=1}^p p(x_j | y = k).$$

Naïve Bayes Prediction

$$\hat{y} = \arg \max_{k \in \{1, 2, \dots, K\}} p(y = k) \prod_{j=1}^p p(x_j | y = k).$$

Estimating prior: $p(y=k)$

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

Estimating likelihood: $p(x_j=v \mid y=k)$

$$\theta_{k,j,v} = \frac{N_{k,j,v} + 1}{N_k + |f_j|}$$

Data Structure idea

(tennis example)

Day	Outlook	Temperature	Humidity	Wind	PlayTennis (y)
x_1	Sunny	Hot	High	Weak	No
x_2	Sunny	Hot	High	Strong	No
x_3	Overcast	Hot	High	Weak	Yes
x_4	Rain	Mild	High	Weak	Yes
x_5	Rain	Cool	Normal	Weak	Yes
x_6	Rain	Cool	Normal	Strong	No
x_7	Overcast	Cool	Normal	Strong	Yes
x_8	Sunny	Mild	High	Weak	No
x_9	Sunny	Cool	Normal	Weak	Yes
x_{10}	Rain	Mild	Normal	Weak	Yes
x_{11}	Sunny	Mild	Normal	Strong	Yes
x_{12}	Overcast	Mild	High	Strong	Yes
x_{13}	Overcast	Hot	Normal	Weak	Yes
x_{14}	Rain	Mild	High	Strong	No

Data Structure idea

(tennis example)

Condition on $y=\text{No}$

Day	Outlook	Temperature	Humidity	Wind	PlayTennis (y)
x_1	Sunny	Hot	High	Weak	No
x_2	Sunny	Hot	High	Strong	No
x_3	Overcast	Hot	High	Weak	Yes
x_4	Rain	Mild	High	Weak	Yes
x_5	Rain	Cool	Normal	Weak	Yes
x_6	Rain	Cool	Normal	Strong	No
x_7	Overcast	Cool	Normal	Strong	Yes
x_8	Sunny	Mild	High	Weak	No
x_9	Sunny	Cool	Normal	Weak	Yes
x_{10}	Rain	Mild	Normal	Weak	Yes
x_{11}	Sunny	Mild	Normal	Strong	Yes
x_{12}	Overcast	Mild	High	Strong	Yes
x_{13}	Overcast	Hot	Normal	Weak	Yes
x_{14}	Rain	Mild	High	Strong	No

Data Structure idea

(tennis example)

y=No (0)

outlook	Sunny:	Overcast:	Rain:
temperature	Cool:	Mild:	Hot:
humidity	Normal:	High:	
wind	Weak:	Strong:	

y=Yes (1)

outlook	Sunny:	Overcast:	Rain:
temperature	Cool:	Mild:	Hot:
humidity	Normal:	High:	
wind	Weak:	Strong:	

$Y = \text{No} (0)$

outlook	Sunny: $\frac{3+1}{5+3}$	overcast: $\frac{0+1}{5+3}$	rain: $\frac{2+1}{5+3}$
:	$\frac{1}{5+3}$	$\frac{1}{5+3}$	$\frac{1}{5+3}$
temp	:	:	:
humidity	$\frac{1}{5+2}$	$\frac{1}{5+2}$	$\frac{1}{5+2}$
wind	$\frac{1}{5+2}$	$\frac{1}{5+2}$	$\frac{1}{5+2}$

Laplace counts

$\log_likelihood[0][\text{"outlook"}][\text{"rain"}]$
 \uparrow list

key is feature name
 value: dictionary

$\{ \text{"outlook"} : \{ \text{"sun"} : \frac{1}{2}, \text{"overcast"} : \frac{1}{8}, \text{"rain"} : \frac{3}{8} \}, \text{"temp"} : \dots \}$

keys are feature values
 values: likelihood

$\theta_{k,j|v_i}$

From previous class meeting

For each method/approach, is X continuous or discrete? What about y ?

- Linear regression
- Polynomial regression
- Decision trees/stumps
- ROC curve as an evaluation metric
- Naïve bayes
- Logistic regression
- Entropy and information gain
- PCA

Think about offline!

Midterm Practice Exam

Questions: 4,5,7,8

If time: 6,9,10

On your own for ~15 min

4, 5, 7, 8

(skip 6)
9, 10

lab

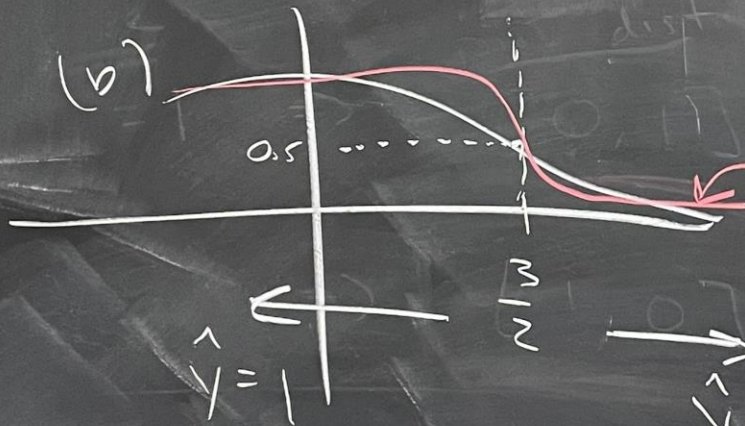
(4)

$$\textcircled{4} \quad \overbrace{3 \cdot 1 - 2x}^{w \cdot x} \geq 0 \Rightarrow$$

$$(a) \quad -2x \geq -3$$

$$\boxed{x \leq \frac{3}{2}}$$

Predict 1
 $\hat{y} = 1$



$$(c) \quad x_{\text{test}} = 2$$

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

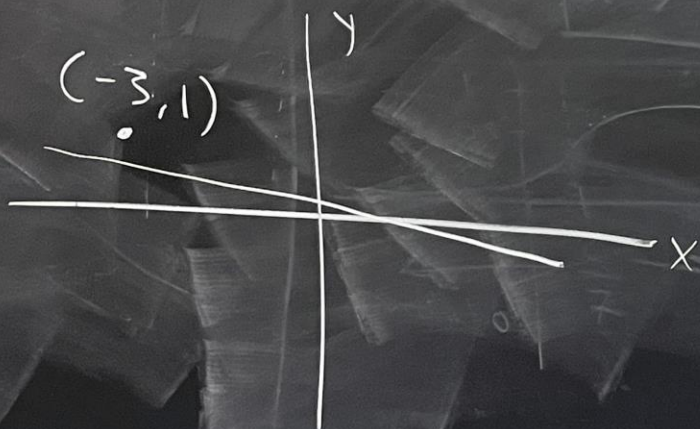
$$(d) \quad 6 - 4x \geq 0$$

$$\Rightarrow \boxed{x \leq \frac{3}{2}}$$

$$(5) \quad \vec{w} \leftarrow \vec{w} - \eta (h_{\vec{w}}(\vec{x}) - y) \vec{x}$$

$$\vec{w} \leftarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.01 \left(\frac{1}{2} - 1 \right) \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

$$\vec{w} \leftarrow \begin{bmatrix} 0.005 \\ -0.015 \end{bmatrix}$$



$$h_{\vec{w}} \left(\begin{bmatrix} 1 \\ -3 \end{bmatrix} \right) = \frac{1}{1 + e^{-\vec{w} \cdot \vec{x}}} = \frac{1}{1 + e^0} = \frac{1}{2}$$

$$\vec{w} \cdot \vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -3 \end{bmatrix} = 0 \cdot 1 + 0 \cdot (-3) = 0$$

after train

$$\vec{w} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

(7)

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

Posterior

$$p(x_1, x_2, x_3 | y) = \frac{p(x_1, x_2, x_3)}{N}$$

likelihood for $p=3$

$$\stackrel{\text{Bayes}}{=} \frac{1}{1 + e^{-\vec{w} \cdot \vec{x}}} = \frac{1}{1 + e^0} = \frac{1}{1 + 1} = \frac{1}{2} \text{ always}$$

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -3 \end{bmatrix} = 0 \cdot 1 + 0 \cdot (-3)$$

$$\textcircled{7} \quad \underbrace{p(\bar{y} | \bar{x})}_{\text{Posterior}} = \frac{\underbrace{p(y)}_{\text{prior}} \underbrace{p(\bar{x} | y)}_{\text{likelihood}}}{\underbrace{p(\bar{x})}_{\text{evidence}}} \quad \left. \vphantom{\frac{p(y)p(\bar{x}|y)}{p(\bar{x})}} \right\} \text{Bayesian model}$$

$$\underbrace{p(x_1, x_2, x_3 | y)}_{\text{likelihood for } p=3} \xrightarrow{\text{Naive Bayes assumption}} p(x_1 | y) p(x_2 | y) p(x_3 | y)$$

$$p(x_2=v|y=k) = \frac{p(x_2=v, y=k)}{p(y=k)} \approx \frac{N_{k,2,v} + 1}{N_k + |f_2|}$$

$$(8.) \quad p(y=0) = \frac{3+1}{5+2} = \frac{4}{7}$$

$$p(y=1) = \frac{2+1}{5+2} = \frac{3}{7}$$

test
 $\vec{x} = [1, 0]$

$$\bullet p(y=0|\vec{x}) \approx p(y=0) p(f_1=1|y=0) p(f_2=0|y=0) / p(\vec{x})$$

$$= \frac{4}{7} \cdot \frac{1+1}{3+3} \cdot \frac{0+1}{3+5} = \frac{4}{7} \cdot \frac{2}{6} \cdot \frac{1}{8}$$

$$\bullet p(y=1|\vec{x}) \approx \frac{6}{245} = 0.0245$$

$$\boxed{\vec{x}=1}$$

$$= \sqrt{\frac{1}{42}} = 0.0238$$

$$\frac{+1}{|f_2|}$$

f_1	f_2	y
3	A	0
2	B	1
1	C	0
2	E	0
1	A	1

(7)

$$\frac{1}{42} = 0.0238$$