

CS 360: Machine Learning

Prof. Sara Mathieson

Fall 2019



- Midterm 1 take-home due **TODAY at 6pm!**
 - Cite sources outside our class notes/textbooks
 - Hand-in under my door (KINSC L302)
- Office hours **today 12:30-1:30pm**
 - Feel free to make an appointment to talk about material or anything else about the course
- **Reading quiz Thursday**
 - <http://cs229.stanford.edu/notes/cs229-notes1.pdf> pg 16-19
- **Lab 5 due October 22** (Tuesday after fall break)
 - Code reviews in lab on Thurs

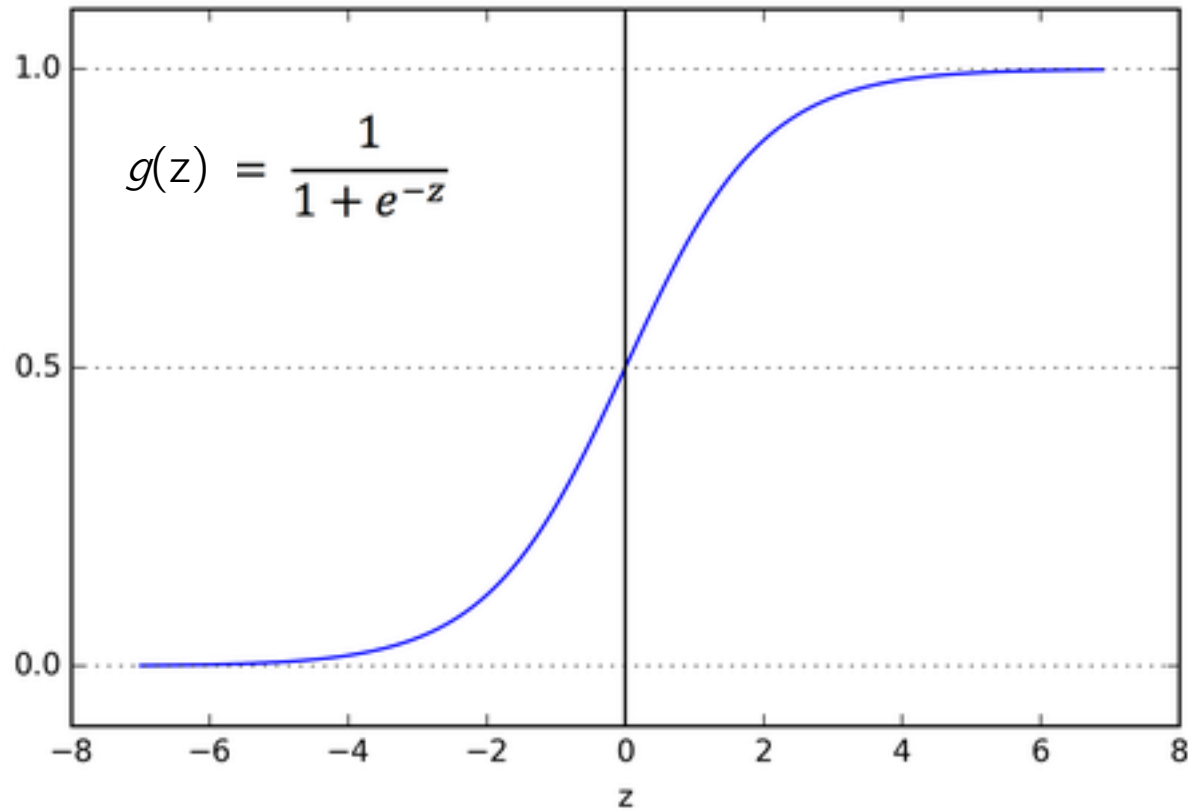
Outline for October 8

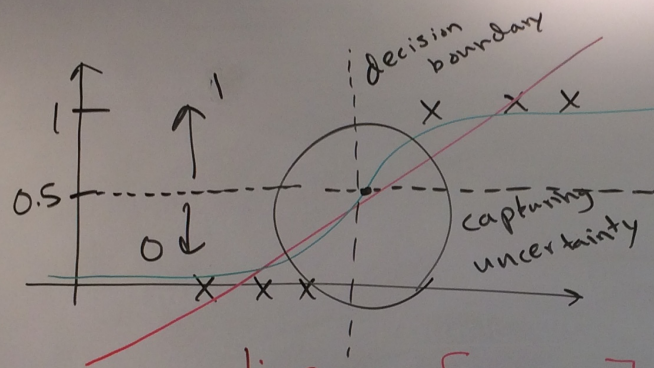
- Logistic regression decision boundaries
- Likelihood functions (Handout 10)
- Logistic regression cost function
- SGD for logistic regression
- Thursday:
 - Multi-class logistic regression
 - Begin: evaluation metrics

Outline for October 8

- Logistic regression decision boundaries
- Likelihood functions (Handout 10)
- Logistic regression cost function
- SGD for logistic regression
- Thursday:
 - Multi-class logistic regression
 - Begin: evaluation metrics

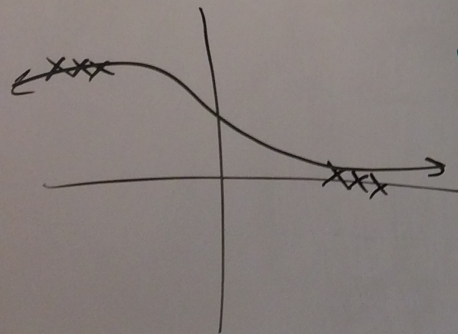
Logistic (sigmoid) function





linear: $[-\infty, \infty] \rightarrow [-\infty, \infty]$

logistic: $[-\infty, \infty] \rightarrow [0, 1]$
classification!



$$g(z) = \frac{1}{1 + e^{-z}}$$

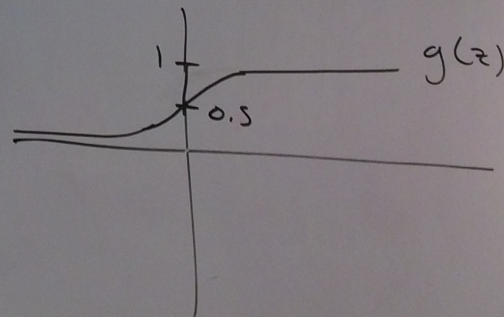
think about
as a probability

logistic
function

idea

$$h_{\vec{w}}(\vec{x}) = p_{\vec{w}}(y=1 | \vec{x})$$

$$= g(\vec{w} \cdot \vec{x})$$



Decision Boundary

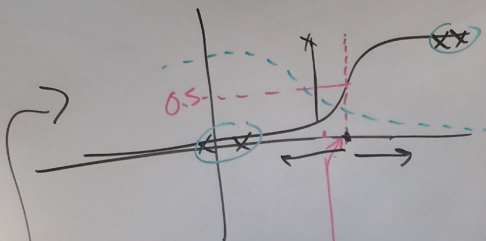
$$\text{set } g(\vec{w} \cdot \vec{x}) = \frac{1}{2}$$

$$\frac{1}{1 + e^{-\vec{w} \cdot \vec{x}}} = \frac{1}{2}$$

$$\Rightarrow 2 = 1 + e^{-\vec{w} \cdot \vec{x}}$$

$$1 = e^{-\vec{w} \cdot \vec{x}}$$

$$\Rightarrow \vec{w} \cdot \vec{x} = 0$$



$$w_0 + w_1 x = 0$$

$$x = -\frac{w_0}{w_1}$$

prediction

$$\text{if } \vec{w} \cdot \vec{x} > 0 \Rightarrow \hat{y} = 1$$

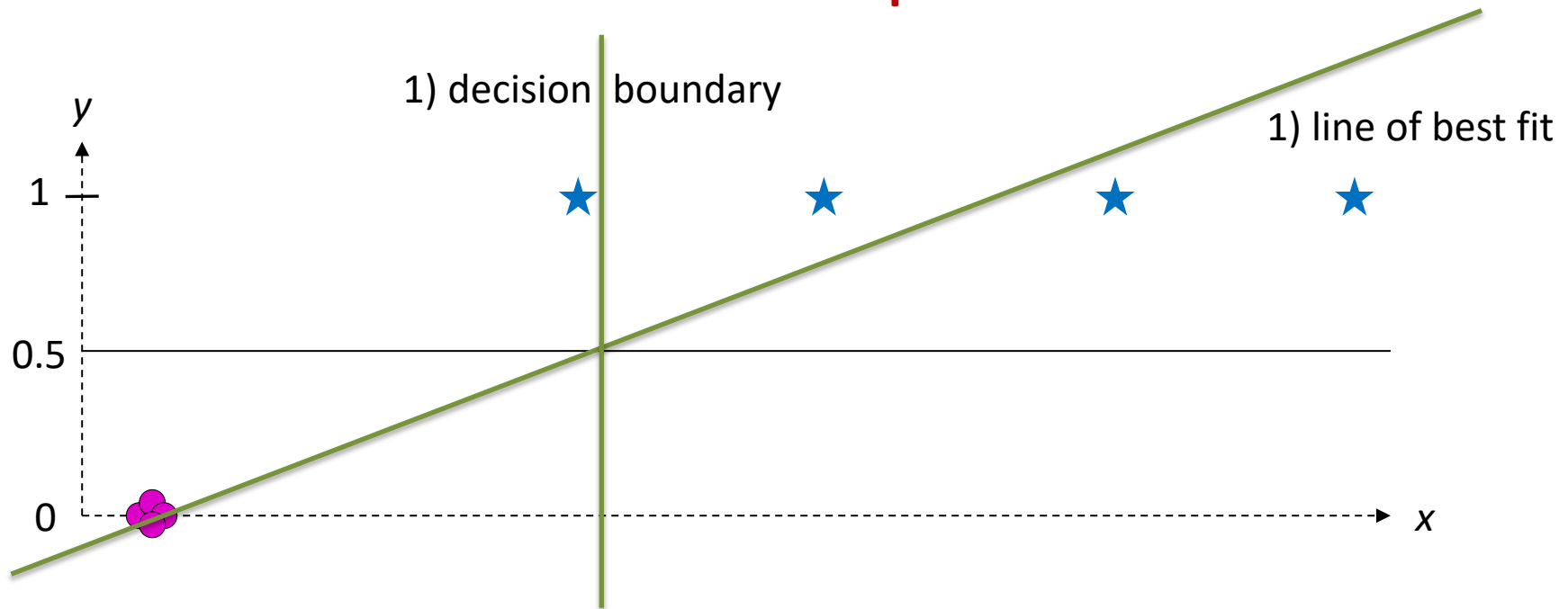
$$\text{if } \vec{w} \cdot \vec{x} \leq 0 \Rightarrow \hat{y} = 0$$

Extra Example



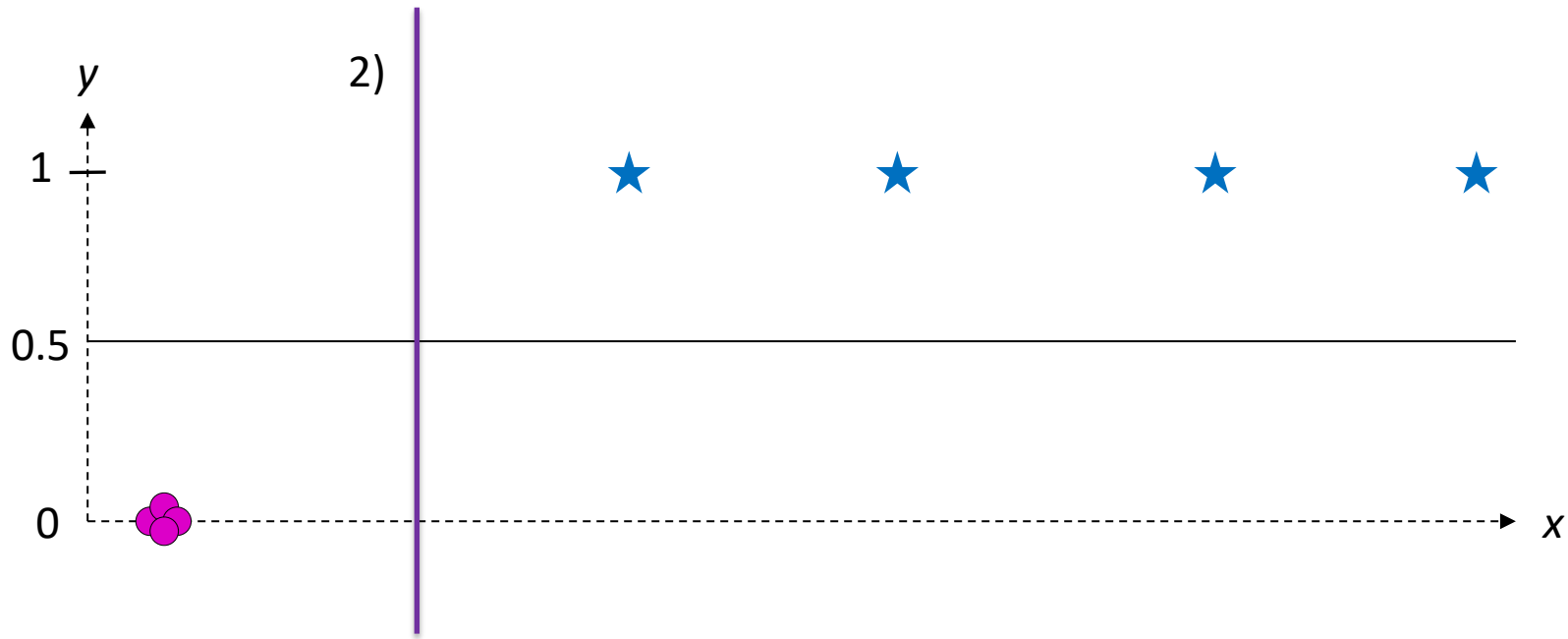
- 1) What line of best fit would be produced by **linear regression**? (roughly)
- 2) What linear decision boundary would be produced by **logistic regression**? (roughly)

Extra Example



- 1) What line of best fit would be produced by **linear regression**? (roughly)
- 2) What linear decision boundary would be produced by **logistic regression**? (roughly)

Extra Example



- 1) What line of best fit would be produced by **linear regression**? (roughly)
- 2) What linear decision boundary would be produced by **logistic regression**? (roughly)

Outline for October 8

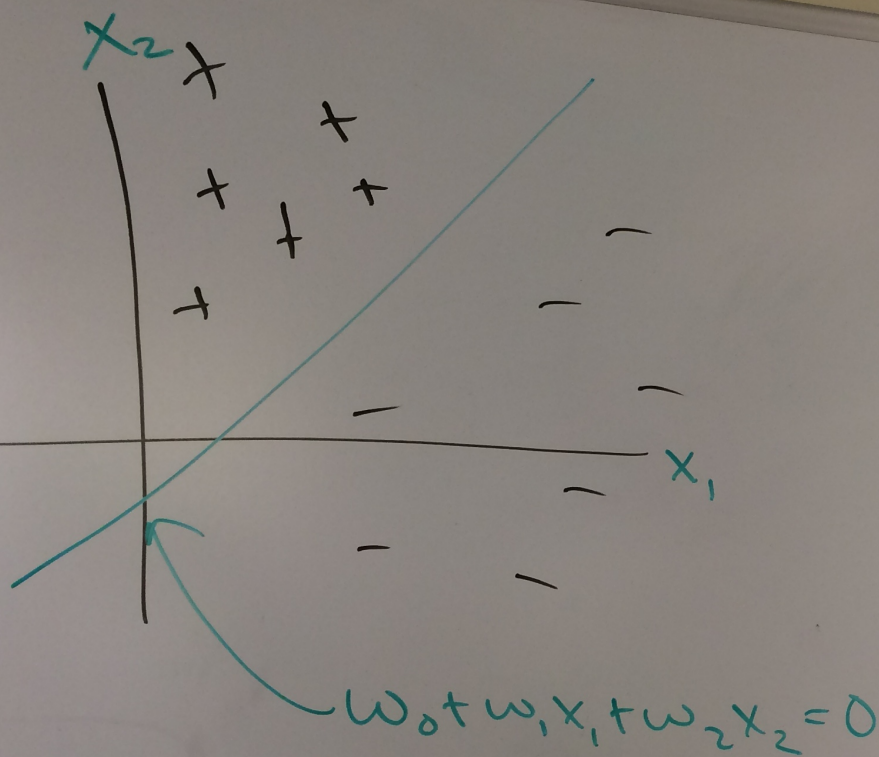
- Logistic regression decision boundaries
- **Likelihood functions (Handout 10)**
- Logistic regression cost function
- SGD for logistic regression

- Thursday:
 - Multi-class logistic regression
 - Begin: evaluation metrics

How to find \vec{w} ?

idea likelihood function $p(w|x,y)$

$$L(\vec{w}) = \prod_{i=1}^n \underbrace{h_{\vec{w}}(\vec{x}_i)^{y_i}}_{\text{prob of label 1}} \underbrace{(1 - h_{\vec{w}}(\vec{x}_i))^{1-y_i}}_{\text{prob of label 0}}$$



$$x_2 = \frac{-w_1x_1 - w_0}{w_2}$$

Aside to Likelihoods

(flip a coin)

$$P(\text{coin H}) = p$$

$$P(\text{coin T}) = 1-p$$

$$\vec{y} = [0, 0, 1, 1, 0, 1, 0, 1, 0, 1]$$

$$L(p) = (1-p)(1-p)pp(1-p) \dots (1-p)$$

$$L(p) = \prod_{i=1}^n p^{y_i} (1-p)^{1-y_i}$$

n coin flips

only one term
chosen for each
flip

HANDOUT 10

Goal: maximize the likelihood

products are difficult (derivative)

log likelihood

$$\ell(p) = \sum_{i=1}^n y_i \log p + \sum_{i=1}^n (1-y_i) \log(1-p)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

mean/average

$$\ell(p) = \log(p) \cdot n\bar{y} + \log(1-p) (n - n\bar{y})$$

$$f(x) = \log x$$

$$f'(x) = \frac{1}{x}$$

take derivative wrt p & set to zero

$$\ell'(p) = \left[\frac{n\bar{y}}{p} - \frac{n - n\bar{y}}{1-p} \right] (p)(1-p) = 0$$

← "inside function"

$$\Rightarrow \hat{p} = \bar{y}$$

Outline for October 8

- Logistic regression decision boundaries
- Likelihood functions (Handout 10)
- Logistic regression cost function
- SGD for logistic regression
- Thursday:
 - Multi-class logistic regression
 - Begin: evaluation metrics

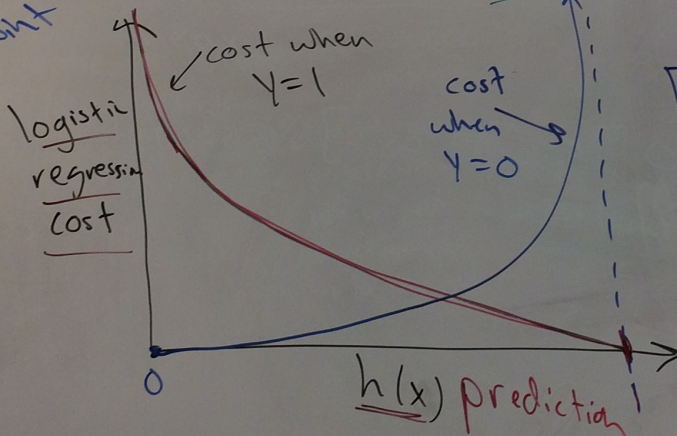
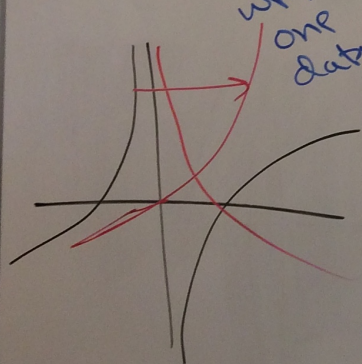
$$\log \quad \ell(\vec{w}) = \sum_{i=1}^n \underbrace{y_i \log(h(\vec{x}_i))}_{y=1} + \underbrace{(1-y_i) \log(1-h(\vec{x}_i))}_{y=0} \quad \text{cost function}$$

$$\Rightarrow J(\vec{w}) = -\ell(\vec{w})$$

minimize cost

$$J(\vec{w}) = \begin{cases} -y \log h(\vec{x}) = -\log h(\vec{x}) & \text{if } \underline{y=1} \\ -(1-y) \log(1-h(\vec{x})) = -\log(1-h(\vec{x})) & \text{if } \underline{y=0} \end{cases}$$

really nice cost function!



useful fact

$$g'(z) = g(z)(1-g(z))$$

$$h_{\vec{w}}(\vec{x}_i) = g(\vec{w} \cdot \vec{x}_i)$$

Exercise!

SGD take derivative

$$\nabla_{\vec{x}_i} J(\vec{w}) = - \left[\frac{y_i}{h_{\vec{w}}(\vec{x}_i)} - \frac{1-y_i}{1-h_{\vec{w}}(\vec{x}_i)} \right] \nabla_{\vec{w}} h_{\vec{w}}(\vec{x}_i)$$

$$= - \left[\frac{y_i}{h_{\vec{w}}(\vec{x}_i)} - \frac{1-y_i}{1-h_{\vec{w}}(\vec{x}_i)} \right] \underbrace{h_{\vec{w}}(\vec{x}_i)(1-h_{\vec{w}}(\vec{x}_i))}_{\text{chain rule}} \vec{x}_i$$

$$= - \left[y_i(1-h_{\vec{w}}(\vec{x}_i)) - (1-y_i)h_{\vec{w}}(\vec{x}_i) \right] \vec{x}_i$$

$$= - \left[y_i - y_i h_{\vec{w}}(\vec{x}_i) - h_{\vec{w}}(\vec{x}_i) + y_i h_{\vec{w}}(\vec{x}_i) \right] \vec{x}_i$$

$$\vec{w} \leftarrow \vec{w} - \alpha \nabla_{\vec{x}_i} J(\vec{w})$$

How to find \vec{w} ?

Idea likelihood function $p(w|x_i)$

come back...

$$L(\vec{w}) = \prod_{i=1}^n \underbrace{h_{\vec{w}}(\vec{x}_i)}_{\text{prob of label 1}}^{y_i} \underbrace{(1-h_{\vec{w}}(\vec{x}_i))}_{\text{prob of label 0}}^{1-y_i}$$

chain rule

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

Same as linear regression!

SGD

$$\vec{w} \leftarrow \vec{w}$$

$$h_{\vec{w}}(\vec{x}_i)$$

fit a logist

SGD

logistic
function

$$\bar{y} = \frac{4}{10} = \frac{\#1's}{n}$$

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

$$\hat{p} = 0.4$$

$$h_{\vec{w}}(\vec{x}_i) = \frac{1}{1 + e^{-\vec{w} \cdot \vec{x}_i}}$$

fit a
logistic function

$$= P_{\vec{w}}(y_i = 1 \mid x_i)$$

Aside to Likelihoods (flip a coin)

$$P(\text{coin H}) = p$$

$$P(\text{coin T}) = 1 - p$$

$$\vec{y} = [0, 0, 1, 1, 0, 1, 0, 1, 0, 0]$$

$$L(p) = (1-p)(1-p)pp(1-p) \dots (1-p)$$

$$L(p) = \prod_{i=1}^n p^{y_i} (1-p)^{1-y_i}$$

only one term
chosen for each
flip

n coin flips