CS 360: Machine Learning

Prof. Sara Mathieson

Fall 2020

HAVERFORD
COLLEGE

Admin

- **Midterm 1** due tonight
  - Grace period (24 hours)

- Office hours **TODAY 5-6:30pm**

- **Reading (pg 14-17)**
  - http://cs229.stanford.edu/notes2020fall/notes2020fall/cs229-notes1.pdf

- **Lab 5 posted today, due next Tues**
  - Small pair exercise in lab Thurs
  - Will check in with everyone on Thurs

Outline for October 13

- Logistic regression decision boundaries
- Likelihood functions (Handout 8)
- Logistic regression cost function
- SGD for logistic regression
- Practice problems (Handout 9)

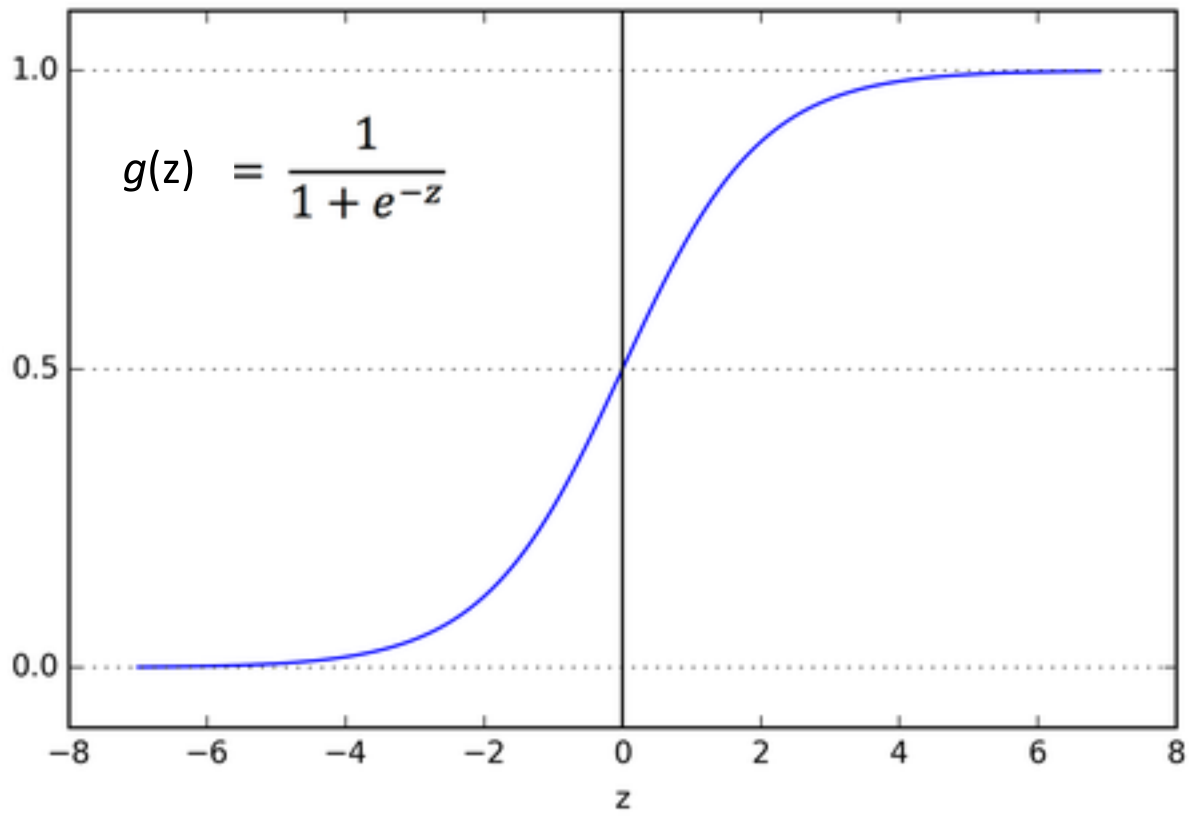
- Friday:
  - Multi-class logistic regression
  - Begin: evaluation metrics

Outline for October 13

- Logistic regression decision boundaries
- Likelihood functions (Handout 8)
- Logistic regression cost function
- SGD for logistic regression
- Practice problems (Handout 9)

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

Logistic (sigmoid) function



Logistic Regression Decision Boundaries

simple $p = 1$

$$\vec{w} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \qquad z = 3 - 2x \implies \underbrace{h_{\vec{w}}(x)}_{p(y=1 \mid x)} = \frac{1}{1 + e^{\underbrace{-(3-2x)}}}$$

linear function

when would I predict 1?

$$\frac{1}{1+e^{-3+2x}} > \frac{1}{2}$$

$$2 > 1 + e^{-3+2x}$$

$$1 > e^{-3+2x}$$

$$0 > -3 + 2x$$

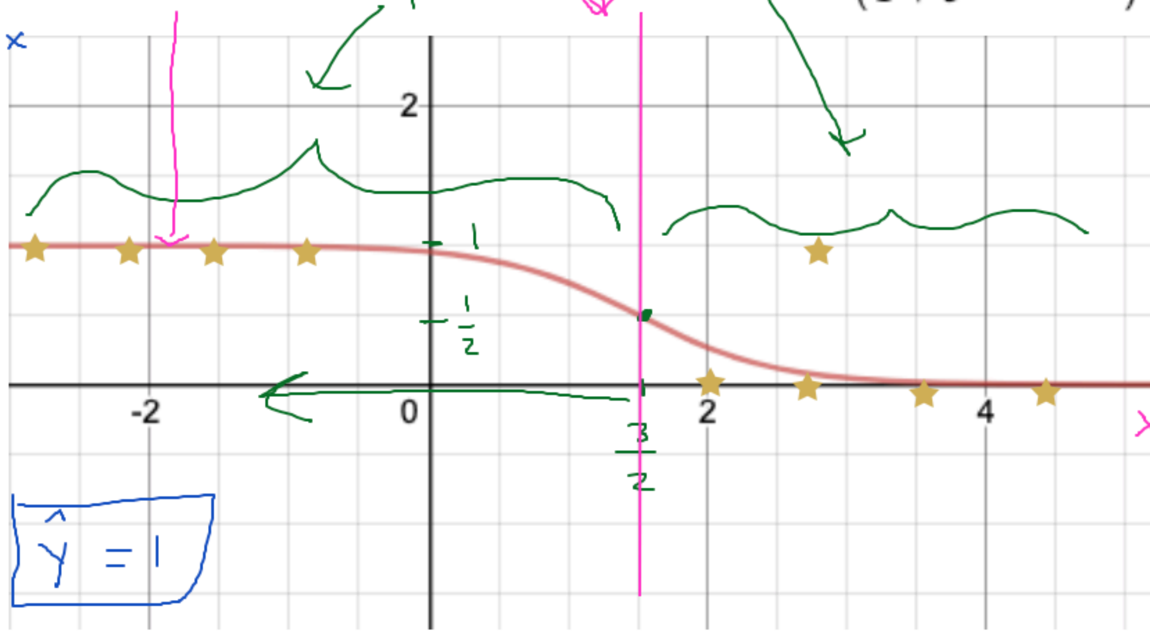
$$3 > 2x$$

decision boundary $\boxed{x < \frac{3}{2}}$ $\Big\}$ $\boxed{\hat{y} = 1}$

$x = \frac{3}{2}$

$h(x)$ $\quad \hat{y} = 1 \quad \hat{y} = 0$

$$\frac{1}{\left(1 + e^{-(3-2x)}\right)}$$

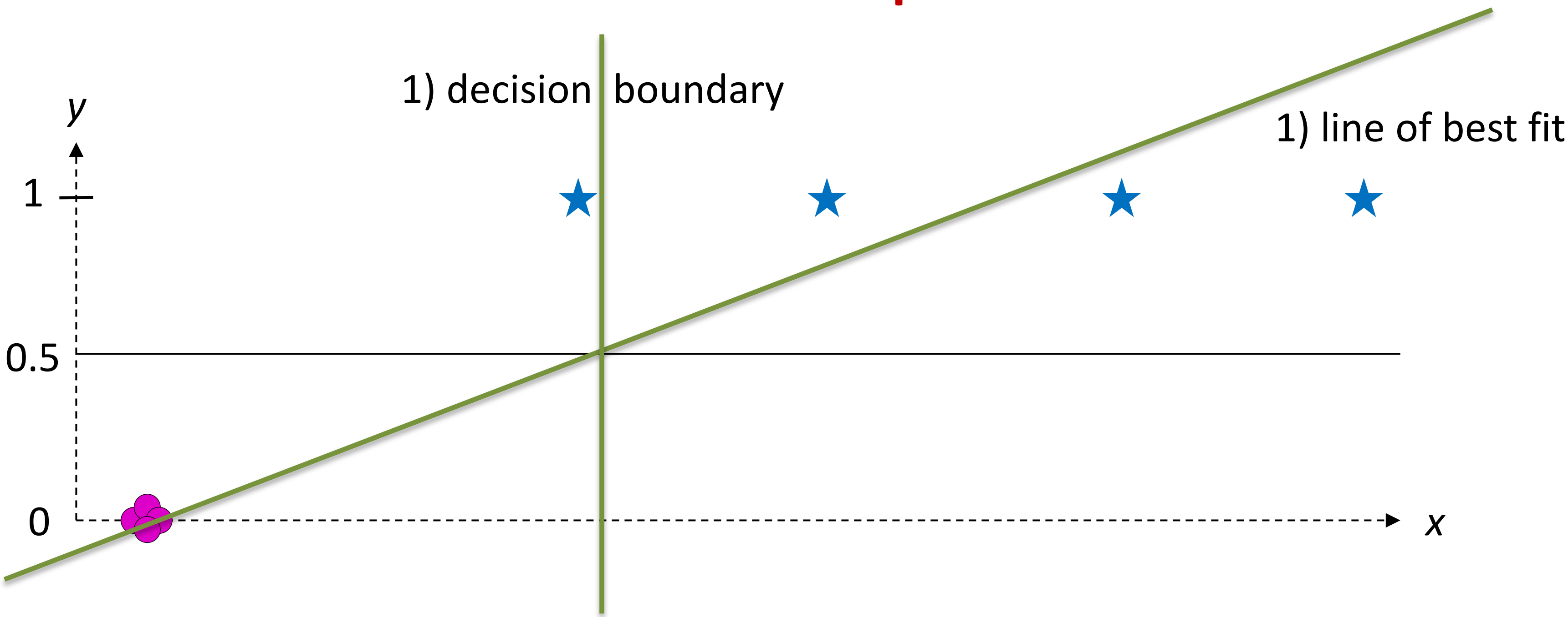


Extra Example

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 13

- Logistic regression decision boundaries
- Likelihood functions (Handout 8)
- Logistic regression cost function
- SGD for logistic regression
- Practice problems (Handout 9)

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

Handout 8

Bernoulli Random Variable. Say we flip a weighted coin n times, and each time the probability of heads (1) is p, so the probability of tails (0) is $(1-p)$. Let y_i be the outcome of flip i. For example, if $n = 10$, we might observe these values:

$p(H) = p$

$p(T) = 1 - p$

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

Goal = find p

In this case, the *likelihood* of p given this observed data is $L(p) = p^4(1-p)^6$, since we observe four 1's and six 0's. In general, we can write the likelihood as

$$L(p) = (1-p)(1-p)p\,p(1-p)p(1-p)p(1-p)(1-p)$$

$$= p^4(1-p)^6$$

Handout 8

Bernoulli Random Variable. Say we flip a weighted coin n times, and each time the probability of heads (1) is p, so the probability of tails (0) is $(1-p)$. Let y_i be the outcome of flip i. For example, if $n = 10$, we might observe these values:

$p(H) = p$

$p(T) = 1 - p$

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

Goal = find p

In this case, the *likelihood* of p given this observed data is $L(p) = p^4(1-p)^6$, since we observe four 1's and six 0's. In general, we can write the likelihood as

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

$$= p^4(1-p)^6$$

in general

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

log

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

Handout 8

Bernoulli Random Variable. Say we flip a weighted coin n times, and each time the probability of heads (1) is p, so the probability of tails (0) is $(1-p)$. Let y_i be the outcome of flip i. For example, if $n = 10$, we might observe these values:

$p(H) = p$

$p(T) = 1-p$

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

$y_1 = 0$

Goal = find p

In this case, the *likelihood* of p given this observed data is $L(p) = p^4(1-p)^6$, since we observe four 1's and six 0's. In general, we can write the likelihood as

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

$$= p^4(1-p)^6$$

$L(p | \vec{y})$

$L(\vec{w} | x, \vec{y})$

in general

pick out $y=1$

picks out $y=0$

cares

$\cancel{p^0}(1-p)^{1-0}$

$= (1-p)$

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

$\underline{\underline{\log}}$

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

Handout 8: solve for p

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

avg

derivative, set to 0

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

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

Handout 8: solve for p

$\underbrace{\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i}_{\text{avg}}$

derivative, set to 0

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

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

$$l'(p) = \frac{n\bar{y}}{p} - \frac{n - n\bar{y}}{1-p} = 0$$

$$\cancel{n}\bar{y}(1-p) = p(\cancel{n} - \cancel{n}\bar{y})$$

$$\bar{y} - \cancel{\bar{y}p} = p - \cancel{p\bar{y}}$$

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

Handout 8: solve for p

$$\underbrace{\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i}_{\text{avg}}$$

derivative, set to 0

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

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

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

Handout

$$\hat{p} = \frac{4}{10}$$

$$\boxed{\hat{p} = \frac{2}{5}}$$

$$\cancel{n}\bar{y}(1-p) = p(\cancel{n} - \cancel{n}\bar{y})$$

$$\bar{y} - \cancel{\bar{y}p} = p - \cancel{p\bar{y}}$$

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

Handout 8: solve for p

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

avg

derivative, set to 0

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

want to maximize

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

known

solve for $\underline{p}$

$$l'(p) = \boxed{\frac{n\bar{y}}{p} - \frac{n-n\bar{y}}{1-p} = 0}$$

6 0's

Handout

$$\hat{p} = \frac{4}{10}$$

1's

n

$$\boxed{\hat{p} = \frac{2}{5}}$$

p(H)

$$\cancel{n}\bar{y}(1-p) = p(\cancel{n} - \cancel{n}\bar{y})$$

$$\bar{y} - \cancel{\bar{y}p} = p - \cancel{p\bar{y}}$$

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

Outline for October 13

- Logistic regression decision boundaries
- Likelihood functions (Handout 8)
- Logistic regression cost function
- SGD for logistic regression
- Practice problems (Handout 9)

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

Logistic Regression Cost Function

max!

$$L(\vec{w}) = \prod_{i=1}^{n} \underbrace{h_{\vec{w}}(\vec{x}_i)}_{p(1)}{}^{y_i \to 1} \underbrace{\left(1 - h_{\vec{w}}(\vec{x}_i)\right)}_{p(0)}{}^{1-y_i \to 0}$$

log likelihood

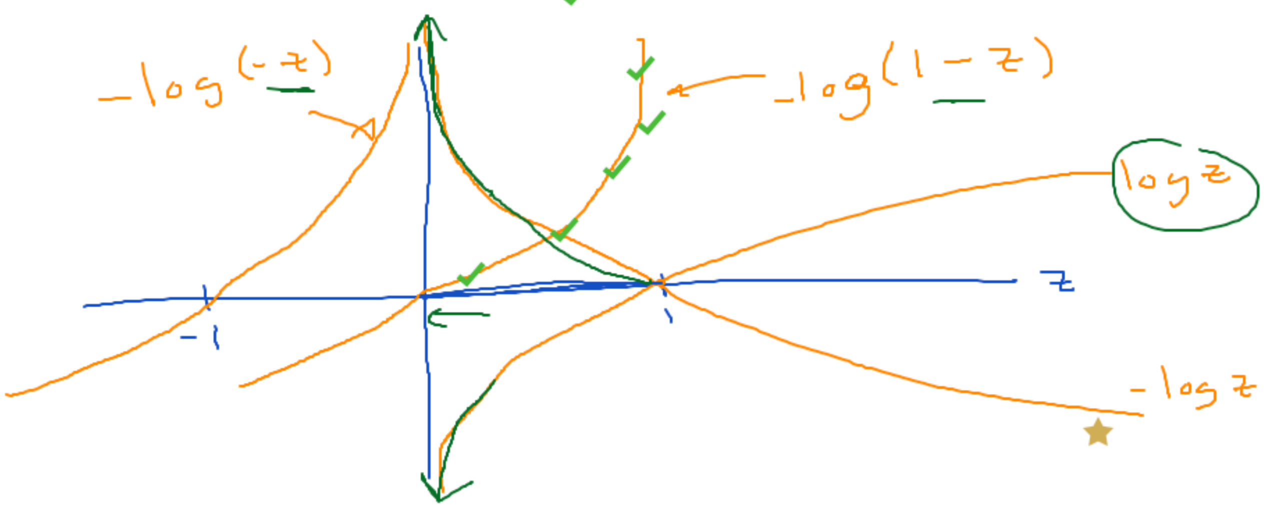
$$\ell(\vec{w}) = \sum_{i=1}^{n} y_i \log h_{\vec{w}}(\vec{x}_i) + \sum_{i=1}^{n} (1-y_i) \log\left(1 - h_{\vec{w}}(\vec{x}_i)\right)$$

cost

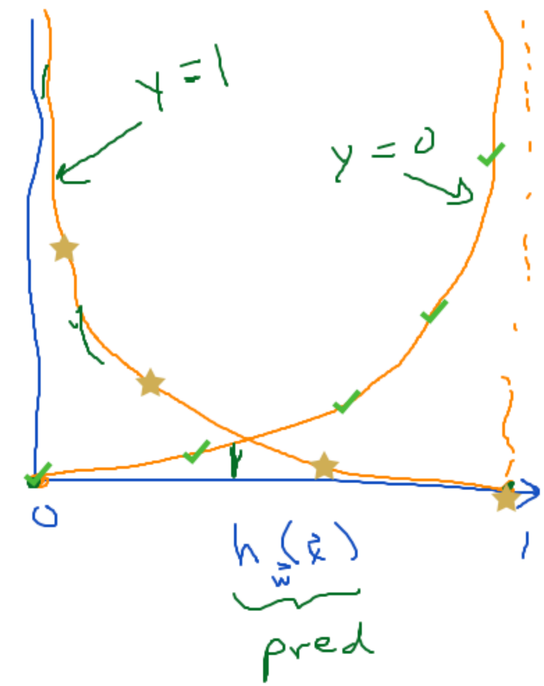
$$\boxed{J(\vec{w}) = -\ell(\vec{w})}$$

minimize cost function

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



cost function



Logistic Regression Cost Function

max!

$$L(\vec{w}) = \prod_{i=1}^{n} \underbrace{h_{\vec{w}}(\vec{x}_i)}_{p(1)}^{\,y_i \to 1} \underbrace{\left(1 - h_{\vec{w}}(\vec{x}_i)\right)}_{p(0)}^{\,1-y_i \to 0}$$

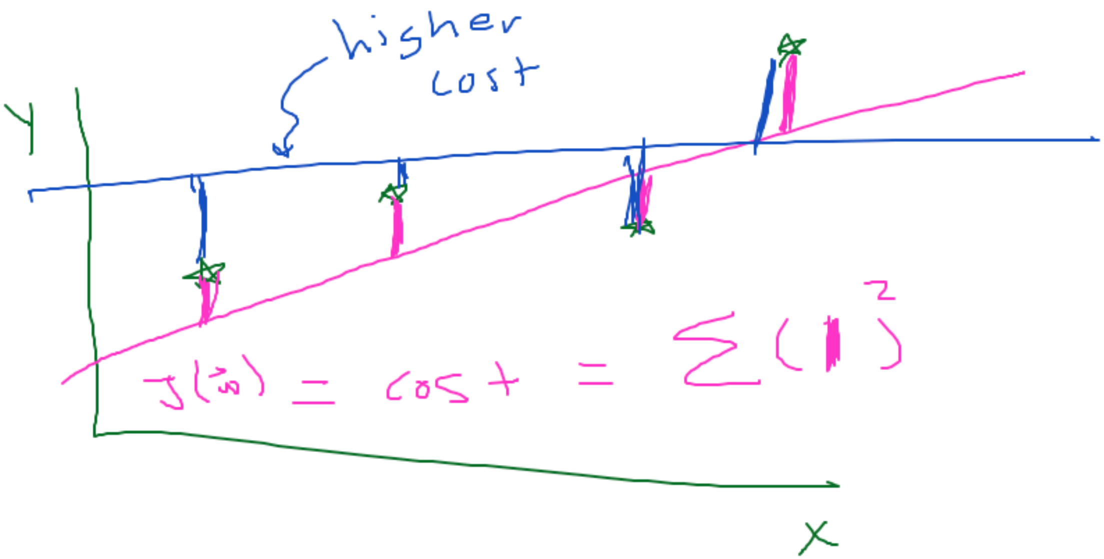
log likelihood

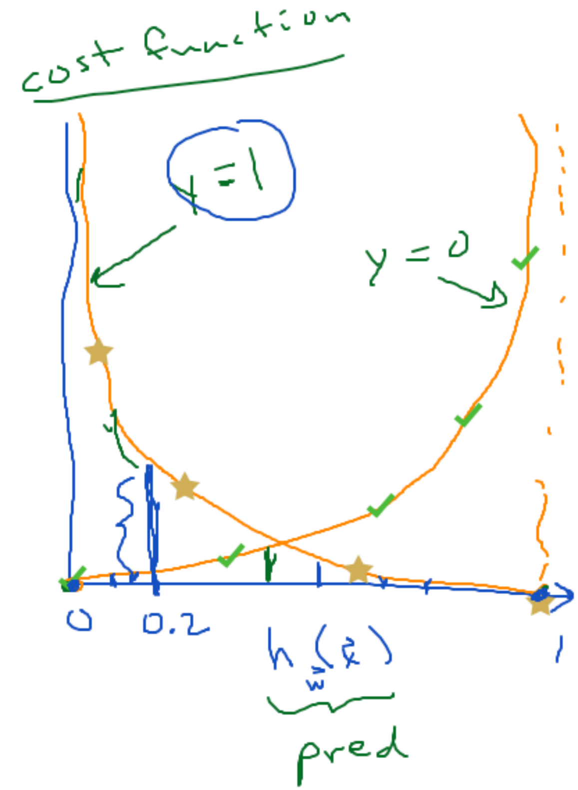
$$\ell(\vec{w}) = \sum_{i=1}^{n} y_i \log h_{\vec{w}}(\vec{x}_i) + \sum_{i=1}^{n} (1-y_i) \log\left(1 - h_{\vec{w}}(\vec{x}_i)\right)$$

cost

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

minimize cost function





Logistic Regression SGD

SGD

for $i = 1 \ldots n$

$$\vec{w} \leftarrow \vec{w} - \alpha \nabla J_{\vec{x}_i}(\vec{w})$$

wrt one data point

$$J = (\vec{w}\cdot\vec{x} - y)^2$$

$$J' = \underbrace{2(\vec{w}\cdot\vec{x} - y)\vec{x}}_{\nabla J}$$

one example

$$\nabla J_{\vec{x}}(\vec{w}) = -\nabla\left(y \log h_{\vec{w}}(\vec{x}) + (1-y)\log(1 - h_{\vec{w}}(\vec{x}))\right)$$

$$= \left(\frac{-y}{h_{\vec{w}}(\vec{x})} + \frac{(1-y)}{1-h_{\vec{w}}(\vec{x})}\right)\nabla h_{\vec{w}}(\vec{x})$$

chain rule

$$= \left[\frac{-y}{h_{\vec{w}}(\vec{x})} + \frac{(1-y)}{1-h_{\vec{w}}(\vec{x})}\right]\underbrace{h_{\vec{w}}(\vec{x})(1-h_{\vec{w}}(\vec{x}))}\vec{x}$$

chain rule

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

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

exercise!

$$z = \vec{w}\cdot\vec{x}$$

$$= \left[-y + \cancel{y\,h(\vec{x})} + h(\vec{x}) - \cancel{y\,h(\vec{x})}\right]\vec{x}$$

$$\boxed{\nabla J(\vec{w}) = (h_{\vec{w}}(\vec{x}) - y)\,\vec{x}}$$

same updates as linear regression

Logistic Regression SGD

SGD

$$\boxed{\vec{w} \leftarrow \vec{w} - \alpha \underbrace{\left(h_{\vec{w}}(\vec{x}_i) - y_i \right) \vec{x}_i}_{\nabla_{\vec{x}_i} J(\vec{w})}}$$

Handout 9

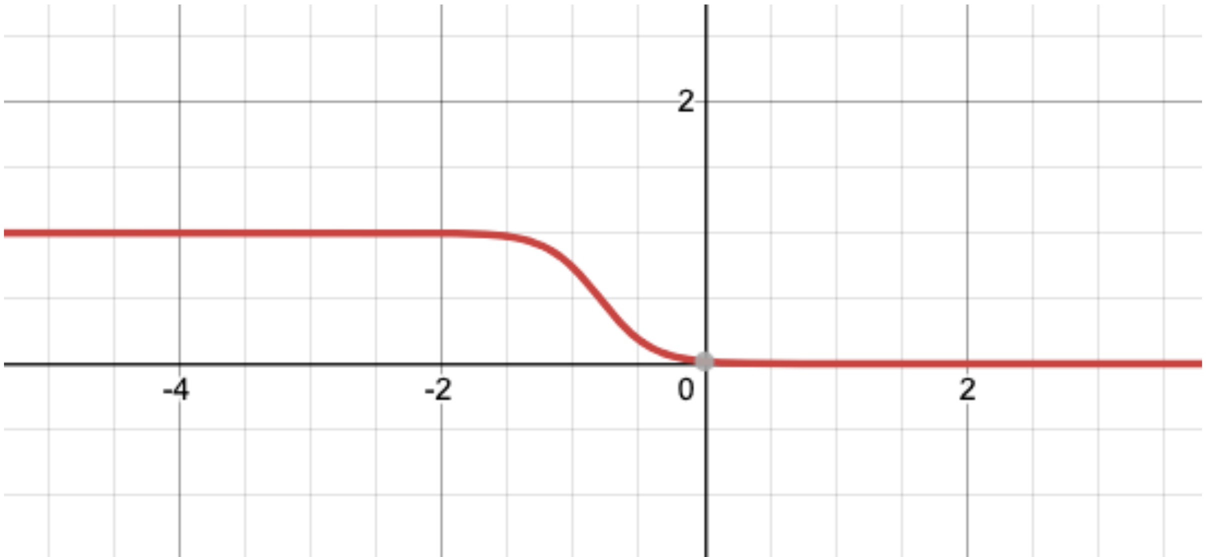
Outline for October 13

- Logistic regression decision boundaries
- Likelihood functions (Handout 8)
- Logistic regression cost function
- SGD for logistic regression
- Practice problems (Handout 9)

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

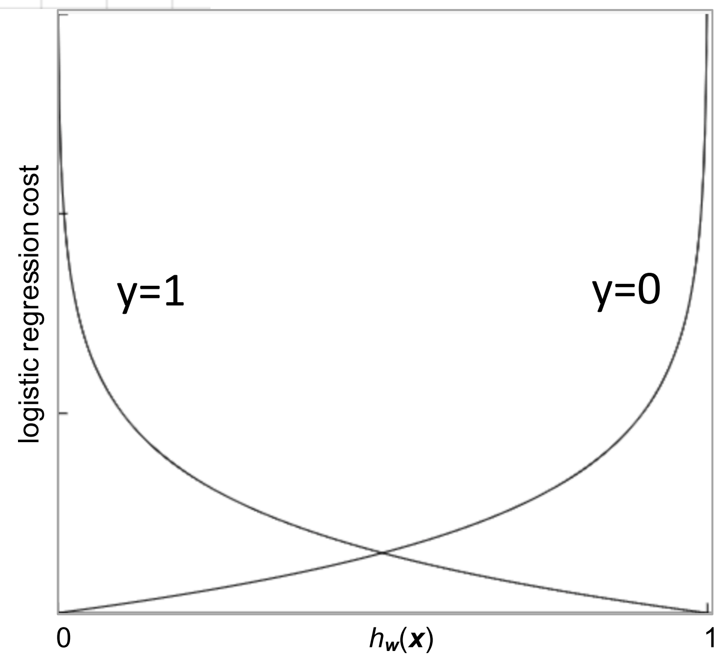
Handout 9

1)



$$\frac{1}{\left(1+e^{-(-4-5x)}\right)}$$

2)



$4 - 5x = 0$

$x = \frac{4}{5}$

$4 - 5x > 0$

$-5x > -4$

$x < +\frac{4}{5}$

$-4 - 5x = 0$

bias

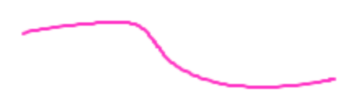
$x = \frac{-4}{5}$

Handout 9

$x \rightarrow \begin{bmatrix} 1 \\ x \end{bmatrix}$

$$\frac{1}{\left(1+e^{-(-4-5x)}\right)}$$

$-w_1$



w_1