CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019
• **Lab 3 due Thursday** (you should be close to done by lab on Wed)
• Midterm 1 Feb 27 (in lab)
  – Let me know ASAP about accommodations
• Office Hours **TODAY 12:30-2pm**
• Choose partner for Lab 4 by Tues night

---

**Research Talk**

Monday, February 18, 2019
11:30-12:30 p.m.
Science Center 240
Michael Wehar
Temple University

“Formal Language Theory and Applications”
Outline for February 18

• Intuition behind logistic regression
• Gradient descent for logistic regression
• Multi-class logistic regression

Soon: statistical evaluation
Outline for February 18

• Intuition behind logistic regression

• Gradient descent for logistic regression

• Multi-class logistic regression

Soon: statistical evaluation
Informal quiz: discuss with a partner

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)
Informal quiz: discuss with a partner

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)
Informal quiz: discuss with a partner

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 February 18

• Intuition behind logistic regression

• Gradient descent for logistic regression

• Multi-class logistic regression

Soon: statistical evaluation
Logistic Regression

Likelihood Function

\[ L(\mathbf{b}) = \prod_{i=1}^{n} \left( h_{\mathbf{b}}(x_i)^{y_i} \left(1-h_{\mathbf{b}}(x_i)\right)^{1-y_i} \right) \]

\[ J(\mathbf{b}) = -\sum_{i=1}^{n} y_i \log h_{\mathbf{b}}(x_i) - \sum_{i=1}^{n} (1-y_i) \log(1-h_{\mathbf{b}}(x_i)) \]

Cost

\[ \text{cost} = \begin{cases} -\log h_{\mathbf{b}}(x_i) & \text{if } y_i = 1 \\ -\log(1-h_{\mathbf{b}}(x_i)) & \text{if } y_i = 0 \end{cases} \]
\[ \text{Cost} = - \sum (1 - y_i) \log (1 - h_b(x_i)) + y_i \log h_b(x_i) \]

\[ \frac{\partial J}{\partial b_j} \]

Graph showing:
- \( y_i = 1 \)
- \( y_i = 0 \)
- \( h_b(x_i) \)
"Inside function" does not include the negative, we already dealt with that!
$$\frac{d J}{d b_j} = \left[ -\frac{y_i}{h_b(x_i)} + \frac{(1-y_i)}{1-h_b(x_i)} \right] \frac{d h_b(x)}{d b_j}$$

$$= \left[ -\frac{y_i}{h_b(x_i)} + \frac{(1-y_i)}{1-h_b(x_i)} \right] h_b(x_i)(1-h_b(x_i))(+x_{ij})$$

$$= -y_i(1-h_b(x_i)) + (1-y_i)(h_b(x_i)) \right] (+x_{ij})$$

$$= [-y_i + y_i h_b(x_i) + h_b(x_i) - y_i h_b(x_i)](+x_{ij})$$

SGD

$$b_j = b_j - \alpha (h_b(x_i) - y_i)x_{ij}$$

Same as linear regression

$$\frac{\partial^2 J}{\partial b_j^2}$$

Cost function graph
Outline for February 18

• Intuition behind logistic regression
• Gradient descent for logistic regression
• Multi-class logistic regression

Soon: statistical evaluation
Multi-class Logistic Regression

- political party
- blood groups

K classes

$y \in \{1, 2, \ldots, K\}$

$$h_b(x) = \frac{1}{1 + e^{-bx}}$$

$$h_b(x) = \frac{e^{bx}}{1 + e^{bx}}$$

$$1 - h_b(x) = \frac{1}{e^{bx} + 1}$$

$y = 1$ weight on class

$y = 0$ weight on class

prob $y = 0$
\[ h_B(x) = \begin{bmatrix}
\frac{p(y=1|x)}{p(y=2|x)}, & \ldots, & \frac{p(y=K|x)}{p(y=2|x)}
\end{bmatrix} \]

Must sum to 1

\[ \text{Normalizer} \]

\[ B = \begin{bmatrix}
\bar{b}^{(1)}, & \bar{b}^{(2)}, & \ldots, & \bar{b}^{(K)}
\end{bmatrix} \]

\[(p+1) \times K\]
Cost function

\[
J_B(\tilde{x}) = \sum_{i=1}^{n} \sum_{k=1}^{K} 1(y_i = k) \log p(y_i = k| \tilde{x}_i)
\]

indicator cost of class \( k \)

if true \( \Rightarrow 1 \)
else \( \Rightarrow 0 \)