Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**

Supervised Learning

Introduction to Data Science Algorithms
Jordan Boyd-Graber and Michael Paul

NOVEMBER 1, 2016

Linear Regression Predictions

dimension	weight
b	1
w_1	2.0
w_2	−1.0
σ	1.0

1. $\mathbf{x}_1 = \{0.0, 0.0\}$; y_1=
2. $\mathbf{x}_2 = \{1.0, 1.0\}$; y_2=
3. $\mathbf{x}_3 = \{.5, 2\}$; y_3=

Linear Regression Predictions

dimension	weight
b	1
w_1	2.0
w_2	-1.0
σ	1.0

1. $\mathbf{x}_1 = \{0.0, 0.0\}$; y_1=1.0
2. $\mathbf{x}_2 = \{1.0, 1.0\}$; y_2=
3. $\mathbf{x}_3 = \{.5, 2\}$; y_3=

Linear Regression Predictions

dimension	weight
b	1
w_1	2.0
w_2	-1.0
σ	1.0

1. $\mathbf{x}_1 = \{0.0, 0.0\}$; y_1=1.0
2. $\mathbf{x}_2 = \{1.0, 1.0\}$; y_2=2.0
3. $\mathbf{x}_3 = \{.5, 2\}$; y_3=

Linear Regression Predictions

dimension	weight
b	1
w_1	2.0
w_2	−1.0
σ	1.0

1. $\mathbf{x}_1 = \{0.0, 0.0\}$; y_1=1.0
2. $\mathbf{x}_2 = \{1.0, 1.0\}$; y_2=2.0
3. $\mathbf{x}_3 = \{.5, 2\}$; y_3=0.0

Probabilities

dimension	weight
w_0	1
w_1	2.0
w_2	-1.0
σ	1.0

$$p(y \mid x) = y \sim N\left(b + \sum_{j=1}^{p} w_j x_j, \sigma^2\right)$$

$$p(y \mid x) = \frac{\exp\left\{-\frac{(y-\hat{y})^2}{2}\right\}}{\sqrt{2\pi}}$$

1. $p(y_1 = 1 \mid \mathbf{x}_1 = \{0.0, 0.0\}) =$
2. $p(y_2 = 3 \mid \mathbf{x}_2 = \{1.0, 1.0\}) =$
3. $p(y_3 = -1 \mid \mathbf{x}_3 = \{.5, 2\}) =$

Probabilities

dimension	weight
w_0	1
w_1	2.0
w_2	-1.0
σ	1.0

$$p(y \mid x) = y \sim N\left(b + \sum_{j=1}^{p} w_j x_j, \sigma^2\right)$$

$$p(y \mid x) = \frac{\exp\left\{-\frac{(y-\hat{y})^2}{2}\right\}}{\sqrt{2\pi}}$$

1. $p(y_1 = 1 \mid \mathbf{x}_1 = \{0.0, 0.0\}) = 0.399$
2. $p(y_2 = 3 \mid \mathbf{x}_2 = \{1.0, 1.0\}) =$
3. $p(y_3 = -1 \mid \mathbf{x}_3 = \{.5, 2\}) =$

Probabilities

dimension	weight
w_0	1
w_1	2.0
w_2	−1.0
σ	1.0

1. $p(y_1 = 1 \,|\, \mathbf{x}_1 = \{0.0, 0.0\}) = 0.399$
2. $p(y_2 = 3 \,|\, \mathbf{x}_2 = \{1.0, 1.0\}) = 0.242$
3. $p(y_3 = -1 \,|\, \mathbf{x}_3 = \{.5, 2\}) =$

$$p(y\,|\,x) = y \sim N\left(b + \sum_{j=1}^{p} w_j x_j, \sigma^2\right)$$

$$p(y\,|\,x) = \frac{\exp\left\{-\frac{(y-\hat{y})^2}{2}\right\}}{\sqrt{2\pi}}$$

Probabilities

dimension	weight
w_0	1
w_1	2.0
w_2	-1.0
σ	1.0

$$p(y \mid x) = y \sim N\left(b + \sum_{j=1}^{p} w_j x_j, \sigma^2\right)$$

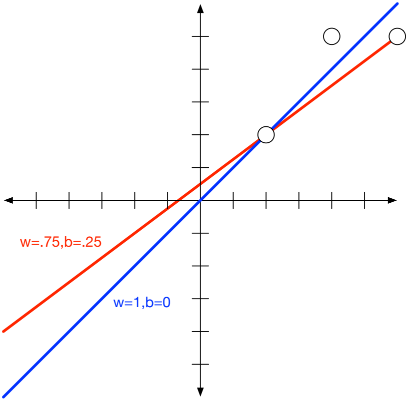
$$p(y \mid x) = \frac{\exp\left\{-\frac{(y-\hat{y})^2}{2}\right\}}{\sqrt{2\pi}}$$

1. $p(y_1 = 1 \mid \mathbf{x}_1 = \{0.0, 0.0\}) = 0.399$
2. $p(y_2 = 3 \mid \mathbf{x}_2 = \{1.0, 1.0\}) = 0.242$
3. $p(y_3 = -1 \mid \mathbf{x}_3 = \{.5, 2\}) = 0.242$

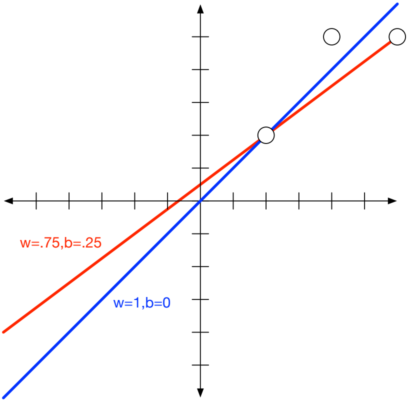
Outline

1. Linear Regression Objective

Consider these points and data

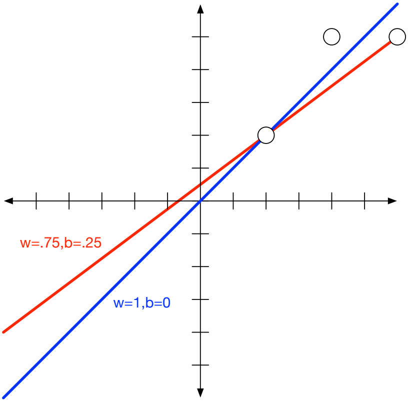


Consider these points and data



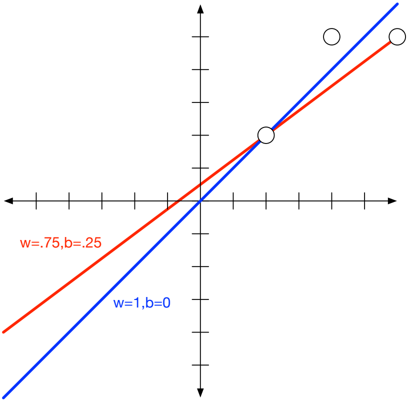
Which is the better OLS solution?

Consider these points and data



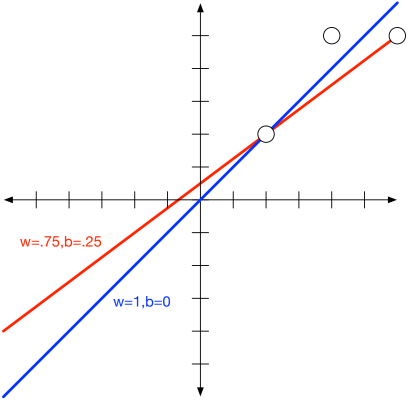
Blue! It has lower RSS.

Consider these points and data



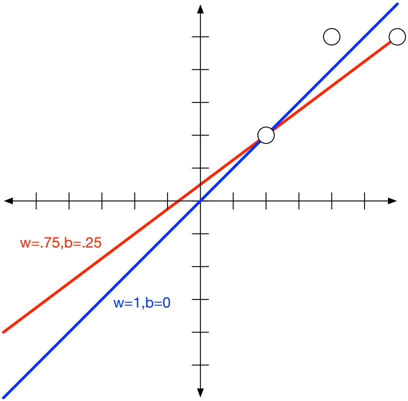
What is the RSS of the better solution?

Consider these points and data



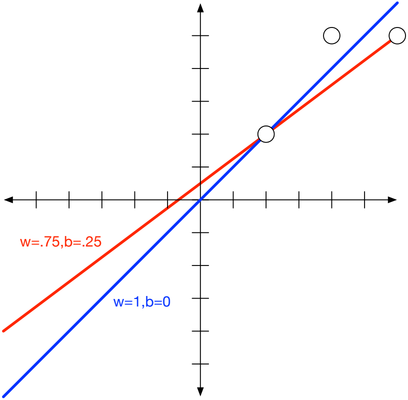
$$\frac{1}{2}\sum_i r_i^2 = \frac{1}{2}\left((1-1)^2+(2.5-2)^2+(2.5-3)^2\right) = \frac{1}{4}$$

Consider these points and data



What is the RSS of the red line?

Consider these points and data



$$\frac{1}{2}\sum_i r_i^2 = \frac{1}{2}\left((1-1)^2 + (2.5-1.75)^2 + (2.5-2.5)^2\right) = \frac{3}{8}$$

Reminder: Logistic Regression

$$P(Y=0|X) = \frac{1}{1+\exp\left[\beta_0 + \sum_i \beta_i X_i\right]} \tag{1}$$

$$P(Y=1|X) = \frac{\exp\left[\beta_0 + \sum_i \beta_i X_i\right]}{1+\exp\left[\beta_0 + \sum_i \beta_i X_i\right]} \tag{2}$$

- Discriminative prediction: $p(y|x)$
- Classification uses: ad placement, spam detection
- What we didn't talk about is how to learn β from data

Logistic Regression: Objective Function

$$\mathscr{L} \equiv \ln p(Y \mid X, \beta) = \sum_j \ln p(y^{(j)} \mid x^{(j)}, \beta) \tag{3}$$

$$= \sum_j y^{(j)} \left(\beta_0 + \sum_i \beta_i x_i^{(j)} \right) - \ln \left[1 + \exp \left(\beta_0 + \sum_i \beta_i x_i^{(j)} \right) \right] \tag{4}$$

Algorithm

1. Initialize a vector B to be all zeros
2. For $t = 1, \ldots, T$
   - For each example $\vec{x}_i, y_i$ and feature j:
     - Compute $\pi_i \equiv \Pr(y_i = 1 \,|\, \vec{x}_i)$
     - Set $\beta[j] = \beta[j]' + \eta(y_i - \pi_i)x_i$
3. Output the parameters $\beta_1, \ldots, \beta_d$.

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle \beta_{bias} = 0, \beta_A = 0, \beta_B = 0, \beta_C = 0, \beta_D = 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

You first see the positive example. First, compute π_1

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\eta = 1.0$.)

You first see the positive example. First, compute π_1

$\pi_1 = \Pr(y_1 = 1 \,|\, \vec{x}_1) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} =$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

You first see the positive example. First, compute π_1

$$\pi_1 = \Pr(y_1 = 1 \,|\, \vec{x}_1) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} = \frac{\exp 0}{\exp 0 + 1} = 0.5$$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

$\pi_1 = 0.5$ What's the update for β_{bias}?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_{bias}?

$$\beta_{bias} = \beta'_{bias} + \eta \cdot (y_1 - \pi_1) \cdot x_{1,bias} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0$$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_{bias}?

$\beta_{bias} = \beta'_{bias} + \eta \cdot (y_1 - \pi_1) \cdot x_{1,bias} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0 \quad = 0.5$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_A?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_A?

$\beta_A = \beta'_A + \eta \cdot (y_1 - \pi_1) \cdot x_{1,A} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 4.0$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_A?

$\beta_A = \beta_A' + \eta \cdot (y_1 - \pi_1) \cdot x_{1,A} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 4.0$ =2.0

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_B?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_B?

$$\beta_B = \beta'_B + \eta \cdot (y_1 - \pi_1) \cdot x_{1,B} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 3.0$$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_B?

$\beta_B = \beta_B' + \eta \cdot (y_1 - \pi_1) \cdot x_{1,B} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 3.0 \quad = 1.5$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_C?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_C?

$\beta_C = \beta'_C + \eta \cdot (y_1 - \pi_1) \cdot x_{1,C} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_C?

$\beta_C = \beta'_C + \eta \cdot (y_1 - \pi_1) \cdot x_{1,C} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0 \quad = 0.5$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_D?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_D?

$$\beta_D = \beta'_D + \eta \cdot (y_1 - \pi_1) \cdot x_{1,D} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 0.0$$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_D?

$\beta_D = \beta'_D + \eta \cdot (y_1 - \pi_1) \cdot x_{1,D} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 0.0$ $=0.0$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

Now you see the negative example. What's π_2?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 =$ **1**

A A A A B B B C

$y_2 =$ **0**

B C C C D D D D

(Assume step size $\eta = 1.0$.)

Now you see the negative example. What's π_2?

$$\pi_2 = \Pr(y_2 = 1 \,|\, \vec{x}_2) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} = \frac{\exp\{.5 + 1.5 + 1.5 + 0\}}{\exp\{.5 + 1.5 + 1.5 + 0\} + 1} =$$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

Now you see the negative example. What's π_2?

$$\pi_2 = \Pr(y_2 = 1 \,|\, \vec{x}_2) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} = \frac{\exp\{.5 + 1.5 + 1.5 + 0\}}{\exp\{.5 + 1.5 + 1.5 + 0\} + 1} = 0.97$$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\eta = 1.0$.)

Now you see the negative example. What's π_2?

$\pi_2 = 0.97$

What's the update for β_{bias}?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_{bias}?

$\beta_{bias} = \beta'_{bias} + \eta \cdot (y_2 - \pi_2) \cdot x_{2,bias} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_{bias}?

$\beta_{bias} = \beta'_{bias} + \eta \cdot (y_2 - \pi_2) \cdot x_{2,bias} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0$ =-0.47

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_A?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_A?

$\beta_A = \beta'_A + \eta \cdot (y_2 - \pi_2) \cdot x_{2,A} = 2.0 + 1.0 \cdot (0.0 - 0.97) \cdot 0.0$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 =$ **1**	$y_2 =$ **0**
A A A A B B B C	B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_A?

$\beta_A = \beta_A' + \eta \cdot (y_2 - \pi_2) \cdot x_{2,A} = 2.0 + 1.0 \cdot (0.0 - 0.97) \cdot 0.0$ =2.0

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_B?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 =$ **1**

A A A A B B B C

$y_2 =$ **0**

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_B?

$\beta_B = \beta'_B + \eta \cdot (y_2 - \pi_2) \cdot x_{2,B} = 1.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_B?

$\beta_B = \beta'_B + \eta \cdot (y_2 - \pi_2) \cdot x_{2,B} = 1.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0$ =0.53

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 =$ **1**

A A A A B B B C

$y_2 =$ **0**

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_C?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 =$ **1**

A A A A B B B C

$y_2 =$ **0**

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_C?

$\beta_C = \beta'_C + \eta \cdot (y_2 - \pi_2) \cdot x_{2,C} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 3.0$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 =$ **1**

A A A A B B B C

$y_2 =$ **0**

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_C?

$\beta_C = \beta'_C + \eta \cdot (y_2 - \pi_2) \cdot x_{2,C} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 3.0$ =-2.41

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_D?

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_D?

$\beta_D = \beta'_D + \eta \cdot (y_2 - \pi_2) \cdot x_{2,D} = 0.0 + 1.0 \cdot (0.0 - 0.97) \cdot 4.0$

Example Documents

$$\beta[j] = \beta[j] + \eta(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\eta = 1.0$.)

What's the update for β_D?

$\beta_D = \beta'_D + \eta \cdot (y_2 - \pi_2) \cdot x_{2,D} = 0.0 + 1.0 \cdot (0.0 - 0.97) \cdot 4.0$ =-3.88

Recap

- Linear Regression
- Logistic Regression
- HW5: Implement SGD for logistic regression