



Stochastic Gradient Descent for Logistic Regression

Natural Language Processing: Jordan
Boyd-Graber
University of Maryland
EXAMPLE

Slides adapted from William Cohen

Reminder: Logistic Regression

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

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

- 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

$$\mathcal{L} \equiv \ln p(Y|X, \beta) = \sum_j \ln p(y^{(j)} | x^{(j)}, \beta) \quad (3)$$

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

Algorithm

1. Initialize a vector B to be all zeros
2. For $t = 1, \dots, T$
 - For each example \vec{x}_i, y_i and feature j :
 - Compute $\pi_i \equiv \Pr(y_i = 1 | \vec{x}_i)$
 - Set $\beta[j] = \beta[j]' + \lambda(y_i - \pi_i)x_i$
3. Output the parameters β_1, \dots, β_d .

Example Documents

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

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

You first see the positive example. First, compute π_1

Example Documents

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

$y_1 = 1$

A A A A B B B C

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

$y_2 = 0$

B C C C D D D D

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

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

$y_1 = 1$

A A A A B B B C

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

$y_2 = 0$

B C C C D D D D

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

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_{bias} ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_{bias} ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_A ?

Example Documents

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

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_A ?

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

Example Documents

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

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_A ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_B ?

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_B ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_B ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_C ?

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_C ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_C ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_D ?

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_D ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_D ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

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

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

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

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

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

$$\pi_2 = 0.97$$

What's the update for β_{bias} ?

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_{bias} ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_{bias} ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_A ?

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_A ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_A ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_B ?

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_B ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_B ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_C ?

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_C ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_C ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_D ?

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_D ?

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

Example Documents

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

$$y_1 = 1$$

A A A A B B B C

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

$$y_2 = 0$$

B C C C D D D D

What's the update for β_D ?

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