Gradients in Logistic Regression and Perceptrons Supplement

CS114B Lab 4

Kenneth Lai

February 17, 2022

Gradients in Logistic Regression

- Cross-entropy loss $L(\hat{y}, y) = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})]$

Gradients in Logistic Regression

- Cross-entropy loss $L(\hat{y}, y) = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})]$
- We want to compute $\frac{\partial L}{\partial w_j}$

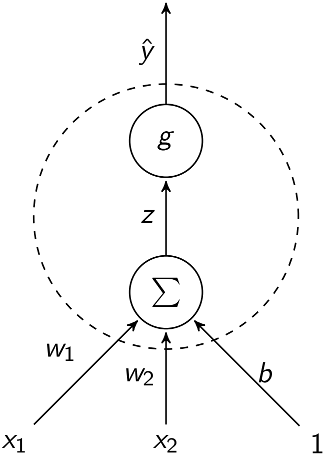
Gradients in Logistic Regression

- Cross-entropy loss $L(\hat{y}, y) = -[y \log \hat{y} + (1 - y) \log(1 - \hat{y})]$
- We want to compute $\frac{\partial L}{\partial w_j}$
- Chain Rule of calculus: $\frac{dy}{dx} = \frac{dy}{dz} \frac{dz}{dx}$

Gradients in Logistic Regression

- Cross-entropy loss $L(\hat{y}, y) = -[y \log \hat{y} + (1-y)\log(1-\hat{y})]$
- We want to compute $\dfrac{\partial L}{\partial w_j}$
- Chain Rule of calculus: $\dfrac{dy}{dx} = \dfrac{dy}{dz}\dfrac{dz}{dx}$
- Looking at the graph: $\dfrac{\partial L}{\partial w_j} = \dfrac{\partial L}{\partial \hat{y}}\dfrac{\partial \hat{y}}{\partial z}\dfrac{\partial z}{\partial w_j}$

Graphical Representation of a Linear Classifier (1)



Gradients in Logistic Regression

- $$\frac{\partial L}{\partial w_j} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z} \frac{\partial z}{\partial w_j}$$

Gradients in Logistic Regression

- $\dfrac{\partial L}{\partial w_j} = \dfrac{\partial L}{\partial \hat{y}} \dfrac{\partial \hat{y}}{\partial z} x_j$
  - $\dfrac{\partial z}{\partial w_j} = x_j$

Gradients in Logistic Regression

- $\dfrac{\partial L}{\partial w_j} = \dfrac{\partial L}{\partial \hat{y}} \hat{y}(1 - \hat{y}) x_j$
  - $\dfrac{\partial z}{\partial w_j} = x_j$
  - For the logistic function: $\dfrac{\partial \hat{y}}{\partial z} = \hat{y}(1 - \hat{y})$

Gradients in Logistic Regression

- $\frac{\partial L}{\partial w_j} = \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})x_j$
  - $\frac{\partial z}{\partial w_j} = x_j$
  - For the logistic function: $\frac{\partial \hat{y}}{\partial z} = \hat{y}(1-\hat{y})$
  - For the cross-entropy loss: $\frac{\partial L}{\partial \hat{y}} = \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}$

Gradients in Logistic Regression

- $\frac{\partial L}{\partial w_j} = \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})x_j = (\hat{y} - y)x_j$
  - $\frac{\partial z}{\partial w_j} = x_j$
  - For the logistic function: $\frac{\partial \hat{y}}{\partial z} = \hat{y}(1-\hat{y})$
  - For the cross-entropy loss: $\frac{\partial L}{\partial \hat{y}} = \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}$

Gradients in Logistic Regression

- $\frac{\partial L}{\partial w_j} = \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})x_j = (\hat{y} - y)x_j$
  - $\frac{\partial z}{\partial w_j} = x_j$
  - For the logistic function: $\frac{\partial \hat{y}}{\partial z} = \hat{y}(1-\hat{y})$
  - For the cross-entropy loss: $\frac{\partial L}{\partial \hat{y}} = \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}$
- $\frac{\partial L}{\partial b} = \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}\hat{y}(1-\hat{y})(1) = \hat{y} - y$
  - $\frac{\partial z}{\partial b} = 1$
  - ...

Gradients in Perceptrons

- Perceptron loss $L(\hat{y}, y) = (\hat{y} - y)z$

Gradients in Perceptrons

- Perceptron loss $L(\hat{y}, y) = (\hat{y} - y)z$
- We want to compute $\frac{\partial L}{\partial w_j}$
- Chain Rule of calculus: $\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$
- Looking at the graph: $\frac{\partial L}{\partial w_j} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial z}\frac{\partial z}{\partial w_j}$

Gradients in Perceptrons

- Perceptron loss $L(\hat{y}, y) = (\hat{y} - y)z$
- We want to compute $\frac{\partial L}{\partial w_j}$
- Chain Rule of calculus: $\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$
- Looking at the graph: $\frac{\partial L}{\partial w_j} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial z}\frac{\partial z}{\partial w_j}$
  - Problem: $\frac{\partial \hat{y}}{\partial z} = 0$ almost everywhere

Gradients in Perceptrons

- Perceptron loss $L(\hat{y}, y) = (\hat{y} - y)z$
- We want to compute $\frac{\partial L}{\partial w_j}$
- Chain Rule of calculus: $\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$
- Looking at the graph: $\frac{\partial L}{\partial w_j} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial z}\frac{\partial z}{\partial w_j}$
  - Problem: $\frac{\partial \hat{y}}{\partial z} = 0$ almost everywhere
    - A change in z does not result in a change in $\hat{y}$ unless you are at the threshold

Gradients in Perceptrons

- Perceptron loss $L(\hat{y}, y) = (\hat{y} - y)z$
- We want to compute $\dfrac{\partial L}{\partial w_j}$
- Chain Rule of calculus: $\dfrac{dy}{dx} = \dfrac{dy}{dz}\dfrac{dz}{dx}$
- Looking at the graph: $\dfrac{\partial L}{\partial w_j} = \dfrac{\partial L}{\partial \hat{y}}\dfrac{\partial \hat{y}}{\partial z}\dfrac{\partial z}{\partial w_j}$
  - Problem: $\dfrac{\partial \hat{y}}{\partial z} = 0$ almost everywhere
    - A change in z does not result in a change in $\hat{y}$ unless you are at the threshold
  - Solution: consider $\dfrac{\partial L}{\partial z}$ directly

Gradients in Perceptrons

- Perceptron loss $L(\hat{y}, y) = (\hat{y} - y)z$
- We want to compute $\dfrac{\partial L}{\partial w_j}$
- Chain Rule of calculus: $\dfrac{dy}{dx} = \dfrac{dy}{dz}\dfrac{dz}{dx}$
- Looking at the graph: $\dfrac{\partial L}{\partial w_j} = \color{blue}{\dfrac{\partial L}{\partial z}}\color{black}{\dfrac{\partial z}{\partial w_j}}$

Gradients in Perceptrons

- $\dfrac{\partial L}{\partial w_j} = \dfrac{\partial L}{\partial z}\dfrac{\partial z}{\partial w_j}$

Gradients in Perceptrons

- $\frac{\partial L}{\partial w_j} = \frac{\partial L}{\partial z} x_j$
  - $\frac{\partial z}{\partial w_j} = x_j$

Gradients in Perceptrons

- $\dfrac{\partial L}{\partial w_j} = (\hat{y} - y)x_j$
  - $\dfrac{\partial z}{\partial w_j} = x_j$
  - For the perceptron loss: $\dfrac{\partial L}{\partial z} = \hat{y} - y$

Gradients in Perceptrons

- $\dfrac{\partial L}{\partial w_j} = (\hat{y} - y)x_j$
  - $\dfrac{\partial z}{\partial w_j} = x_j$
  - For the perceptron loss: $\dfrac{\partial L}{\partial z} = \hat{y} - y$
- $\dfrac{\partial L}{\partial b} = \hat{y} - y$
  - $\dfrac{\partial z}{\partial b} = 1$
  - ...