CS 383: Machine Learning

Prof Adam Poliak

Fall 2024

10/08/2024

Lecture 14

Reading Quiz 5

1. The output of logistic regression is a model that creates:

   (a) a linear decision boundary

   (b) a logistic decision boundary

   (c) no decision boundary

Reading Quiz 5

1. The output of logistic regression is a model that creates:
   (a) a linear decision boundary
   (b) a logistic decision boundary
   (c) no decision boundary

Reading Quiz 5

1. The output of logistic regression is a model that creates:
   - (a) a linear decision boundary
   - (b) a logistic decision boundary
   - (c) no decision boundary
2. We use logistic regression for:
   - (a) classification
   - (b) regression
   - (c) both

Reading Quiz 5

1. The output of logistic regression is a model that creates:
   (a) a linear decision boundary
   (b) a logistic decision boundary
   (c) no decision boundary

2. We use logistic regression for:
   (a) classification
   (b) regression
   (c) both

Reading Quiz 5

1. The output of logistic regression is a model that creates:
   (a) $\boxed{\text{a linear decision boundary}}$
   (b) a logistic decision boundary
   (c) no decision boundary

2. We use logistic regression for:
   (a) $\boxed{\text{classification}}$
   (b) regression
   (c) both

3. Our hypothesis in logistic regression is:

$$h_{\boldsymbol{w}}(\boldsymbol{x}) = p(y = 1|\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{w} \cdot \boldsymbol{x}}}$$

   If $\boldsymbol{w}$ is the zero vector (as it would be when starting SGD), what is the probability $y = 1$?

Reading Quiz 5

1. The output of logistic regression is a model that creates:

   (a) a linear decision boundary

   (b) a logistic decision boundary

   (c) no decision boundary

2. We use logistic regression for:

   (a) classification

   (b) regression

   (c) both

3. Our hypothesis in logistic regression is:

$$h_{\boldsymbol{w}}(\boldsymbol{x}) = p(y = 1|\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{w}\cdot\boldsymbol{x}}}$$

   If $\boldsymbol{w}$ is the zero vector (as it would be when starting SGD), what is the probability $y = 1$?

½

Reading Quiz 5

1. The output of logistic regression is a model that creates:
   (a) a linear decision boundary
   (b) a logistic decision boundary
   (c) no decision boundary

2. We use logistic regression for:
   (a) classification
   (b) regression
   (c) both

3. Our hypothesis in logistic regression is:

$$h_{\boldsymbol{w}}(\boldsymbol{x}) = p(y = 1|\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{w}\cdot\boldsymbol{x}}}$$

   If $\boldsymbol{w}$ is the zero vector (as it would be when starting SGD), what is the probability $y = 1$?

½

4. How did we define the cost function for logistic regression? (Bonus: write down the cost function)
   (a) likelihood
   (b) log likelihood
   (c) negative log likelihood

Reading Quiz 5

1. The output of logistic regression is a model that creates:
   (a) a linear decision boundary
   (b) a logistic decision boundary
   (c) no decision boundary

2. We use logistic regression for:
   (a) classification
   (b) regression
   (c) both

3. Our hypothesis in logistic regression is:

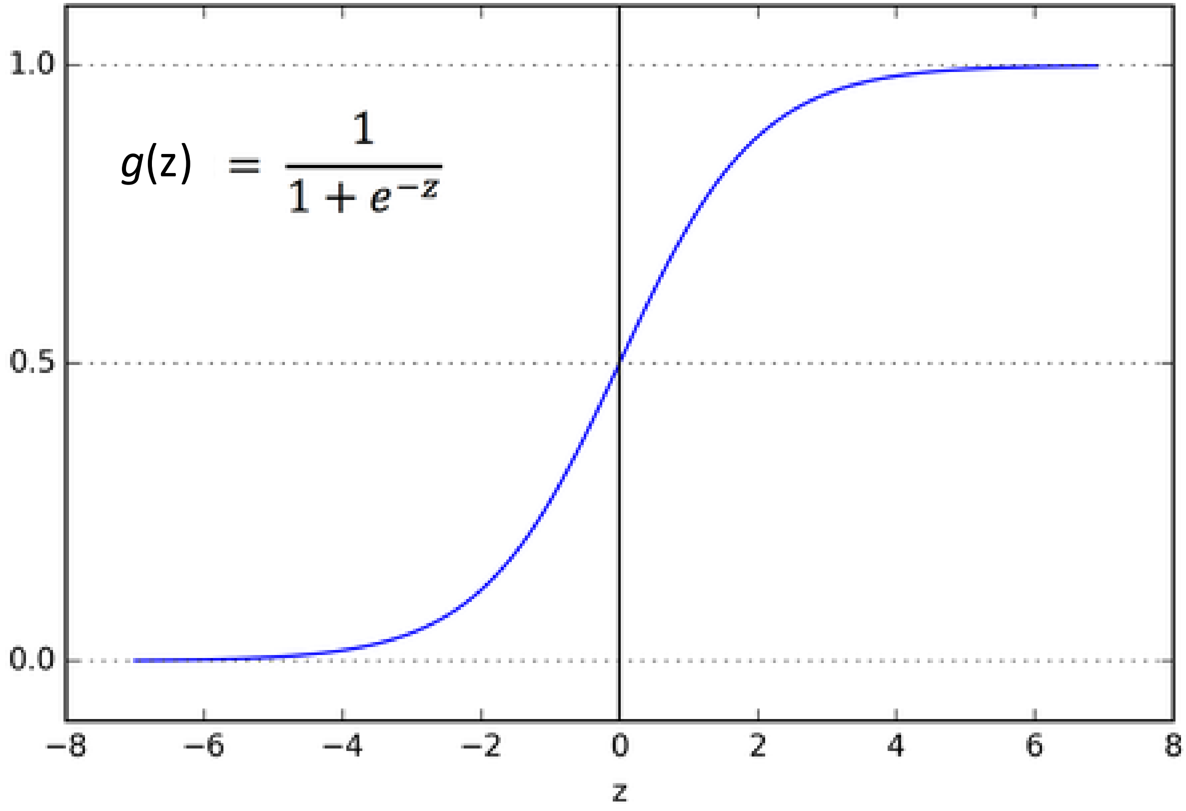
$$h_{\boldsymbol{w}}(\boldsymbol{x}) = p(y = 1|\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{w}\cdot\boldsymbol{x}}}$$

   If $\boldsymbol{w}$ is the zero vector (as it would be when starting SGD), what is the probability $y = 1$?

½

4. How did we define the cost function for logistic regression? (Bonus: write down the cost function)
   (a) likelihood
   (b) log likelihood
   (c) negative log likelihood

Logistic (sigmoid) function



Log-likelihood functions

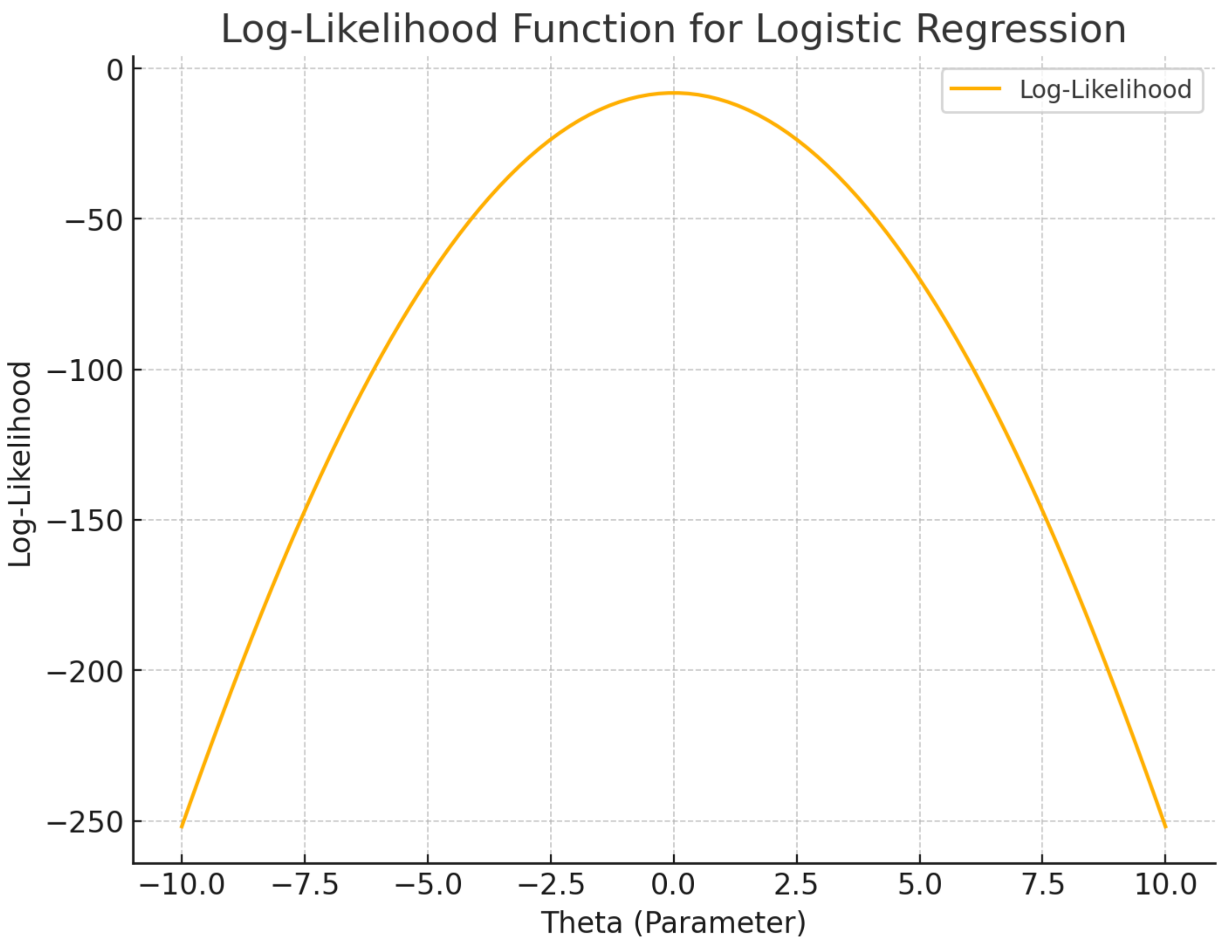
Last class's derivation and handout

Cost/Loss Function for Logistic Regression

$$L(\vec{w}) = \sum_{i}^{n} \left(y_i * \log\left(h_w(x_i)\right) + (1-y)log\left(1-h_w(x_i)\right) \right)$$

We want to minimize our cost function $J(\vec{w})$.

In linear regression we set $J(\vec{w}) = L(\vec{w})$



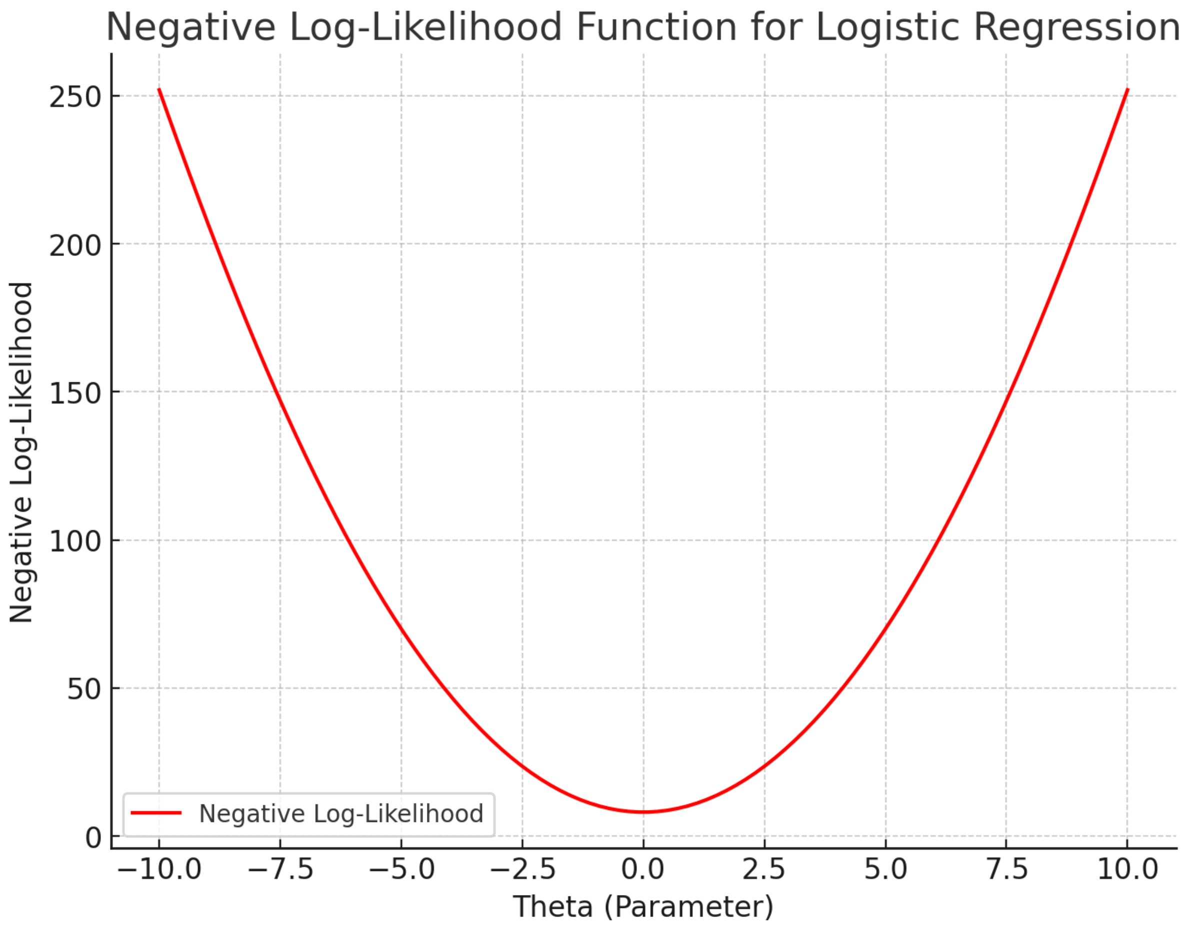
Cost/Loss Function for Logistic Regression

Negative Log Likelihood:

$$J(\vec{w}) = -L(\vec{w})$$

Now we can minimize our cost

No closed form-solution exists



SGD for Logistic Regression

$$\theta \leftarrow \theta \ - \alpha \frac{\partial J(\theta)}{\partial \theta}$$

$$J(\theta) = y * \log\big(h_\theta(x)\big) + (1 - y) * log\big(1 - h_\theta(x)\big)$$

$$h_\theta(x) = \ \frac{1}{1 + e^{-\theta^t x}}$$

Regularization

$$||\vec{w}||_0 = \sum_{j:w_j \neq 0} 1$$

$$||\vec{w}||_1 = \sum_{j=1}^{p} |w_j|$$

$$||\vec{w}||_2 = \sqrt{\sum_{j=1}^{p} w_j^2}$$

L_0 norm

- Number of non-zero entries
- Minimizing L_0 norm is NP hard

L_1 norm

- Sum of magnitude of weights
- Not differentiable

L_2 norm

- Sum of squared weights
- Differentiable

Multi-class prediction