Logistic Regression

Introduction to Data Science Algorithms
Jordan Boyd-Graber and Michael Paul
ABC
Logistic Regression: Objective Function

\[ \ell \equiv \ln p(Y \mid X, \beta) = \sum_j \ln p(y^{(j)} \mid x^{(j)}, \beta) \]

\[ = \sum_j y^{(j)} \left( \beta_0 + \sum_i \beta_i x_i^{(j)} \right) - \ln \left[ 1 + \exp \left( \beta_0 + \sum_i \beta_i x_i^{(j)} \right) \right] \]
Logistic Regression: Objective Function

\[ \ell \equiv \ln p(Y \mid X, \beta) = \sum_j \ln p(y^{(j)} \mid x^{(j)}, \beta) \]

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

(2)

Training data \((y, x)\) are fixed. Objective function is a function of \(\beta\) ... what values of \(\beta\) give a good value.
Convexity

- Convex function
- Doesn’t matter where you start, if you “walk up” objective
Convexity

- Convex function
- Doesn’t matter where you start, if you “walk up” objective
- Gradient!
Gradient Ascent (non-convex)

**Goal**
Optimize log likelihood with respect to variables $\beta$
Gradient Ascent (non-convex)

Goal
Optimize log likelihood with respect to variables $\beta$
Gradient Ascent (non-convex)

Goal
Optimize log likelihood with respect to variables $\beta$

![Graph showing the objective function with respect to the parameter $\theta$. The graph has multiple peaks and valleys, illustrating the non-convex nature of the objective function.](image)
Gradient Ascent (non-convex)

Goal
Optimize log likelihood with respect to variables $\beta$

![Graph showing the optimization process of log likelihood with respect to variables $\beta$. The graph illustrates a non-convex function with multiple local optima and the gradient ascent process moving towards an undiscovered country parameter.]
Gradient Ascent (non-convex)

Goal
Optimize log likelihood with respect to variables $\beta$
Gradient Ascent (non-convex)

Goal

Optimize log likelihood with respect to variables $\beta$
Gradient Ascent (non-convex)

Goal
Optimize log likelihood with respect to variables $\beta$

Diagram:
- Objective vs. Parameter
- Undiscovered Country
- Points 0, 1, and 2 indicating steps in the gradient ascent process.
Gradient Ascent (non-convex)

Goal

Optimize log likelihood with respect to variables $\beta$
Gradient Ascent (non-convex)

Goal
Optimize log likelihood with respect to variables $\beta$
Gradient Ascent (non-convex)

Goal

Optimize log likelihood with respect to variables $\beta$

Luckily, (vanilla) logistic regression is convex