ECE595 / STAT598: Machine Learning I
Lecture 15.3: Logistic Regression 2 - Comparison with Linear Regression

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering
Purdue University

PURDUE UNIVERSITY

Outline

Discriminative Approaches

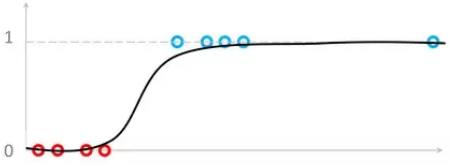
- Lecture 14 Logistic Regression 1
- Lecture 15 Logistic Regression 2

This lecture: Logistic Regression 2

- Gradient Descent
  - Convexity
  - Gradient
  - Regularization
- Connection with Bayes
  - Derivation
  - Interpretation
- Comparison with Linear Regression
  - Is logistic regression better than linear?
  - Case studies

Is Logistic Regression Better than Linear?

Logistic regression on the other hand can handle this outlier with no issue.



Now let's take a closer look at the logistic regression loss function.

$$f(\mathbf{w}) = \sum_{p} \log\left(1 + e^{-y_p \mathbf{x}_p^T \mathbf{w}}\right)$$

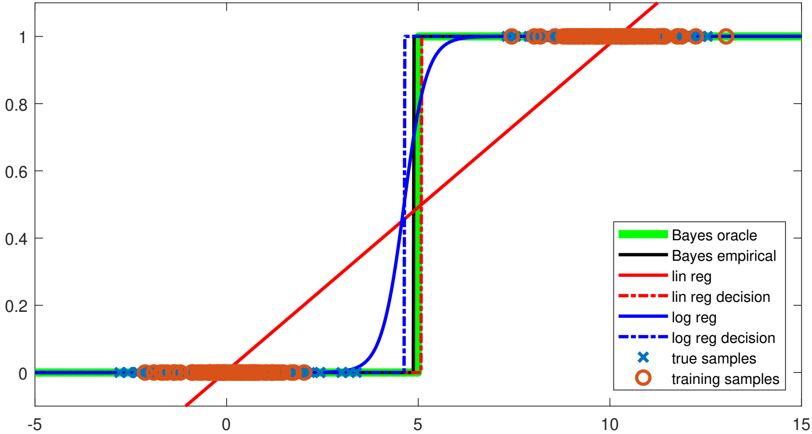
Here, I'm assuming the labels y_p are in $\{-1, +1\}$. Note that this is equivalent

Upvote · 11 Share

- This is taken from the Internet
- Is that true???

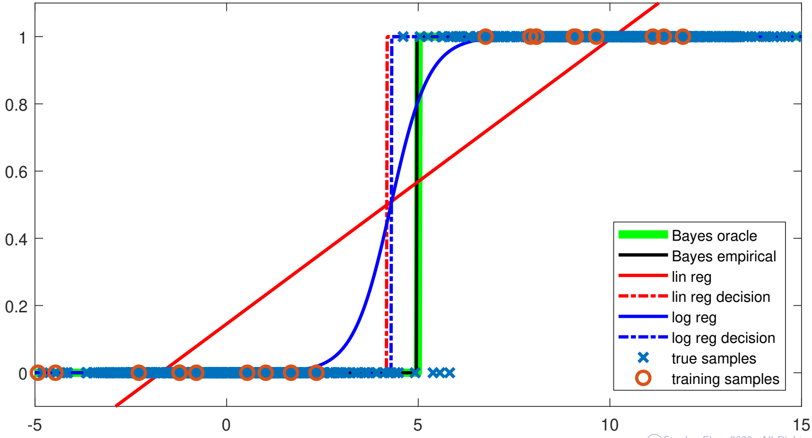
Is Logistic Regression Better than Linear?

- **Scenario 1**: Identical Covariance. Equal Prior. Enough samples.
- $\mathcal{N}(0,1)$ with 100 samples and $\mathcal{N}(10,1)$ with 100 samples.
- Linear and logistic: Not much different.



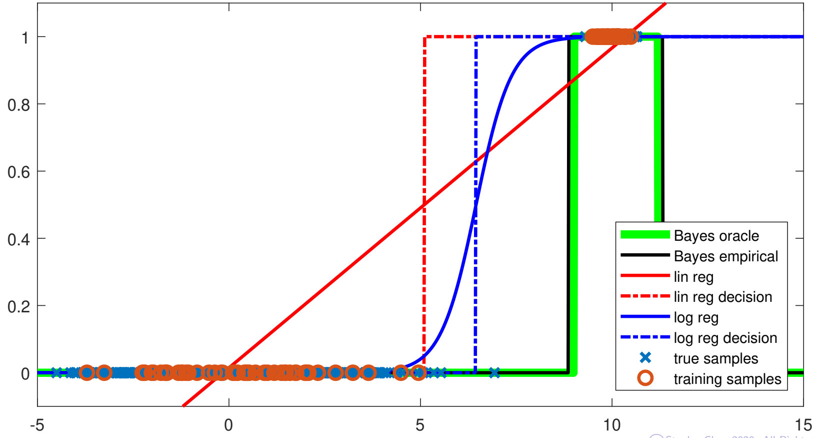
The False Sense of Good Fitting

- **Scenario 2**: Identical Covariance. Equal Prior. Not a lot of samples.
- $\mathcal{N}(0, 2)$ with 10 samples and $\mathcal{N}(10, 2)$ with 10 samples.
- Linear and logistic: Not much different.



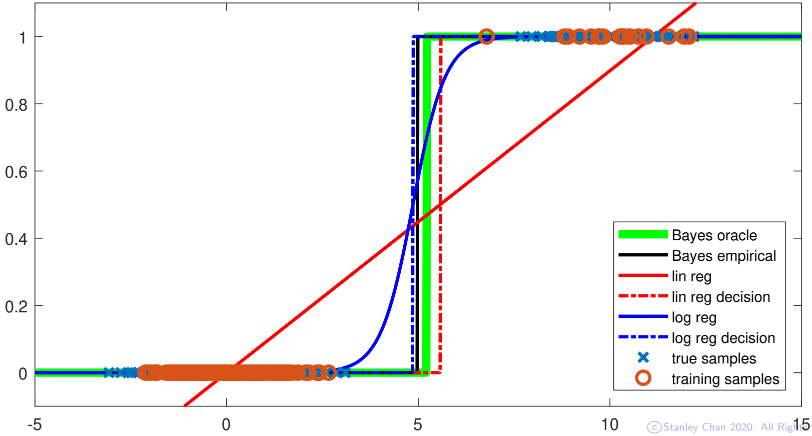
Is Logistic Regression Better than Linear?

- **Scenario 3**: Different Covariance. Equal Prior.
- $\mathcal{N}(0, 2)$ with 50 samples and $\mathcal{N}(10, 0.2)$ with 50 samples.
- Linear and logistic: Equally bad.



Is Logistic Regression Better than Linear?

- **Scenario 4**: Identical Covariance. Unequal Prior.
- Training size proportional to prior: 180 samples and 20 samples.
- $\mathcal{N}(0, 1)$ with $\pi_0 = 0.9$ and $\mathcal{N}(10, 1)$ with $\pi_1 = 0.1$.
- Linear and logistic: Not much different.



So what can we say about Logistic Regression?

- Logistic regression empowers a discriminative method with probabilistic reasonings.
- The hypothesis function is the posterior probability

$$
\begin{aligned}
p(1|\boldsymbol{x}) &= \frac{1}{1+\exp\{-(\boldsymbol{w}^T\boldsymbol{x}+w_0)\}} = h_{\boldsymbol{\theta}}(\boldsymbol{x}) \\
p(0|\boldsymbol{x}) &= \frac{\exp\{-(\boldsymbol{w}^T\boldsymbol{x}+w_0)}{1+\exp\{-(\boldsymbol{w}^T\boldsymbol{x}+w_0)\}} = 1 - h_{\boldsymbol{\theta}}(\boldsymbol{x}),
\end{aligned}
$$

- Logistic is yet another special case of Bayesian
- More or less the same performance as linear regression
- Logistic can give lower training error — which looks better on plots.
- But its generalization is similar to linear regression

Reading List

Logistic Regression (Machine Learning Perspective)

- Chris Bishop's *Pattern Recognition*, Chapter 4.3
- Hastie-Tibshirani-Friedman's *Elements of Statistical Learning*, Chapter 4.4
- Stanford CS 229 Discriminant Algorithms http://cs229.stanford.edu/notes/cs229-notes1.pdf
- CMU Lecture https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf
- Stanford Language Processing https://web.stanford.edu/~jurafsky/slp3/ (Lecture 5)

Logistic Regression (Statistics Perspective)

- Duke Lecture https://www2.stat.duke.edu/courses/Spring13/sta102.001/Lec/Lec20.pdf
- Princeton Lecture https://data.princeton.edu/wws509/notes/c3.pdf