Foundations of Machine Learning
AI2000 and AI5000

FoML-20
Logistic Regression - SGD

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

So far in FoML

- Intro to ML and Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Supervised learning
    a. Linear Regression with basis functions (regularization, model selection)
    b. Bias-Variance Decomposition (Bayesian Regression)
    c. Decision Theory - three broad classification strategies
        - Probabilistic Generative Models - Continuous & discrete data
        - (Linear) Discriminant Functions - least squares solution, Perceptron
        - Probabilistic Discriminative Models - Logistic Regression

Logistic Regression - SGD

Logistic Regression for 2 classes

- Conditional likelihood of the data:

$$p(\mathbf{t}|\mathbf{X},\mathbf{w}) = \prod_{i=1}^{N} p(t_i|\mathbf{x_i},\mathbf{w}) = \prod_{i=1}^{N} {y_i}^{t_i}(1-y_i)^{1-t_i}$$

- The NLL:

$$E(\mathbf{w}) = -\log p(\mathbf{t}|\mathbf{X},\mathbf{w}) = -\left[\sum_{i=1}^{N} t_i \log(y_i) + (1-t_i)\log(1-y_i)\right]$$

Logistic Regression for 2 classes

- SGD for the cross-entropy loss $$E(\mathbf{w}) = -\left[\sum_{i=1}^{N} t_i \log(y_i) + (1 - t_i) \log(1 - y_i)\right]$$

Chain rule of differentiation

Rough

Next
Newton Raphson method