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

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

Data-driven Intelligence
& Learning Lab

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

$$\frac{\partial}{\partial \underline{w}}\left[E(\underline{w})\right] = 0 = -\left[\sum_{i=1}^{N} t_i \frac{1}{\cancel{y_i}} \cancel{y_i}(1-y_i)\underline{\phi}_i^T + (1-t_i)\frac{-1}{\cancel{(1-y_i)}} y_i \cancel{(1-y_i)}\phi_i^T\right]$$

$$= -\left[\sum_{i=1}^{N} t_i(1-y_i)\phi_i^T - (1-t_i)y_i\phi_i^T\right]$$

$$= -\left[\sum_{i=1}^{N}(t_i\phi_i^T - y_i\phi_i^T)\right] = \sum_{i=1}^{N} \underbrace{(y_i - t_i)\phi_i^T}_{\nabla E_i(\underline{w})}$$

$$\phi_i = \underline{\phi}(x_i)_{m\times 1}$$

Chain rule of differentiation

$$\frac{\partial}{\partial x} f[g(x)] = f'[g(x)] \cdot g'(x)$$

very useful for applying Gradient Descent

Rough

Next
Newton Raphson method