Foundations of Machine Learning AI2000 and AI5000

FoML-21
Logistic Regression - Newton Raphson optimization

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 - IRLS

Loss function approximation

- Taylor series expansion (univariate case)

$$E(w + \Delta w) =$$

Loss function approximation

- Taylor series expansion (multivariate case)

$$E(\mathbf{w} + \Delta\mathbf{w}) =$$

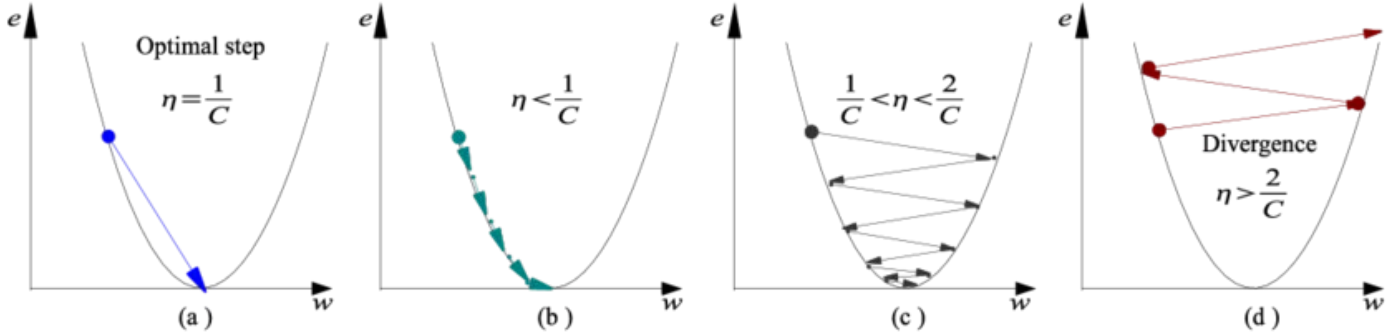
Linear (first-order) approximation

$$E(\mathbf{w} + \Delta\mathbf{w}) \approx E(\mathbf{w}) + \Delta\mathbf{w}^T \nabla E(\mathbf{w})$$

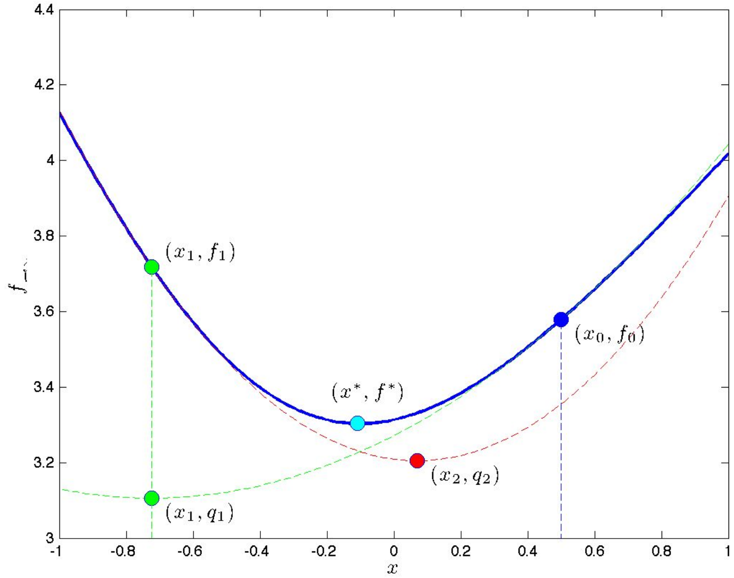
Quadratic (second-order) approximation

$$E(\mathbf{w} + \Delta\mathbf{w}) \approx E(\mathbf{w}) + \Delta\mathbf{w}^T \nabla E(\mathbf{w}) + \frac{1}{2}\Delta\mathbf{w}^T \nabla^2 E(\mathbf{w})\Delta\mathbf{w}$$

Convergence for Quadratic functions



Generic Convex functions



Newton-Raphson Iterative optimization

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \mathbf{H}^{-1}\nabla E(\mathbf{w}^t)$$

$$\nabla E(\mathbf{w})^T = \sum_{n=1}^{N}(y_n - t_n)\phi_n =$$

Newton-Raphson Iterative optimization

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \mathbf{H}^{-1}\nabla E(\mathbf{w}^t)$$

$$\mathbf{H}_{ij} = \frac{\partial^2 E(\mathbf{w}^t)}{\partial \mathbf{w}_i \partial \mathbf{w}_j} = \frac{\partial}{\partial \mathbf{w}_i} \sum_{n=1}^{N} (y_n - t_n)\phi_j(\mathbf{x}_n) =$$

Newton-Raphson Iterative optimization

$$\mathbf{w}^{t+1} = \mathbf{w}^{t} - \mathbf{H}^{-1}\nabla E(\mathbf{w}^{t})$$

$$\begin{aligned}\mathbf{w}^{t+1} &= \mathbf{w}^{t} - (\Phi^T R\Phi)^{-1}\Phi^T(\mathbf{y} - \mathbf{t}) \\ &= (\Phi^T R\Phi)^{-1}\Phi^T R\mathbf{z}\end{aligned}$$

Next
Neural Networks