Foundations of Machine Learning
AI2000 and AI5000

FoML-19
Probabilistic Discriminative Models - Logistic Regression

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

Classification Strategies

- Discriminant functions
  - Direct functions of i/p to target $t = y(\mathbf{x}, \mathbf{w})$
- Probabilistic Discriminant models
  - Posterior class probabilities $p(C_k/\mathbf{x})$
- Probabilistic generative models
  - Class-conditional models $p(\mathbf{x}/C_k)$
  - Prior class probabilities $p(C_k)$

Logistic Regression for 2 classes

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)^T \qquad \mathbf{t} = (t_1, t_2, \ldots, t_N)^T \qquad t_n \in \mathcal{C}_1, \mathcal{C}_2 = \{0, 1\}$$

Logistic Regression for 2 classes

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)^T \qquad \mathbf{t} = (t_1, t_2, \ldots, t_N)^T \qquad t_n \in \mathcal{C}_1, \mathcal{C}_2 = \{0, 1\}$$

- Basis functions: $\phi = \phi(\mathbf{x}) =$

- Probabilistic Discriminative Models: posteriors $p(\mathcal{C}_k|\phi)$ are nonlinear functions over a linear function of ϕ

$$p(\mathcal{C}_k|\phi, \mathbf{w}) =$$

Logistic Regression for 2 classes

$$p(\mathcal{C}_1|\phi, \mathbf{w}) = y(\phi) = \sigma(\mathbf{w}^T\phi)$$

$$p(\mathcal{C}_2|\phi, \mathbf{w}) =$$

$$p(t|\phi, \mathbf{w}) = y(\phi) =$$

Logistic Regression for 2 classes

- Probabilistic discriminative models need less parameters than the Generative models
  - Gaussian class conditional densities $p(\mathbf{x}|\mathcal{C}_k) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\right\}$
  - Class priors $p(C_k)$

Logistic Regression for 2 classes

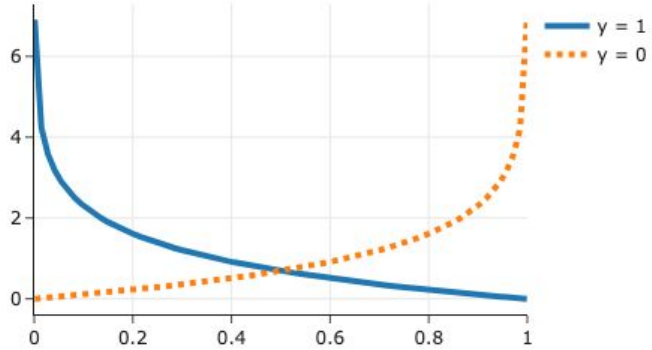
- Conditional likelihood of the data:

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}) =$$

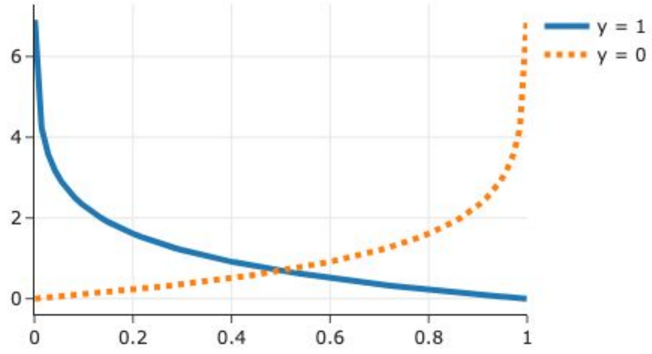
- Minimizing the NLL:

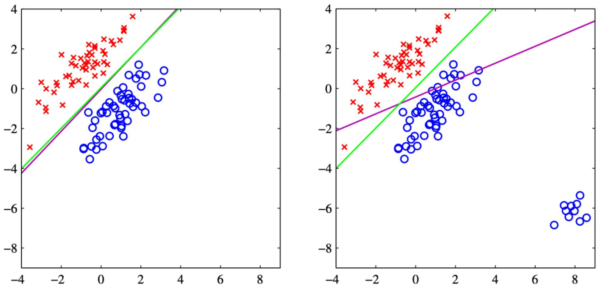
$$E(\mathbf{w}) =$$

Cross-entropy loss



Cross-entropy loss (robust to outliers)





Next
Learning the parameters of Logistic Regression