Foundations of Machine Learning
AI2000 and AI5000

FoML-16
Least Squares for 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
        - Discriminant Functions

Least Squares for Classification

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

DiL
Data-driven Intelligence
& Learning Lab

Least Squares for Classification

- Consider K classes
- Each class ‘k’ has its own linear model $y_k(\mathbf{x}) = w_k^T \mathbf{x} + w_{k0}$

Least Squares for Classification

- Shorter notation $y(\mathbf{x}) = \widetilde{\mathbf{W}}^T \tilde{\mathbf{x}}$

$\widetilde{\mathbf{W}} =$

$\tilde{\mathbf{x}} =$

$y(\mathbf{x}) =$

Assign x to C_k, where

Least Squares for Classification

- Data matrix
- Target matrix

Use regression (sum of squares) error function

$$E_D(\widetilde{\mathbf{W}}) =$$

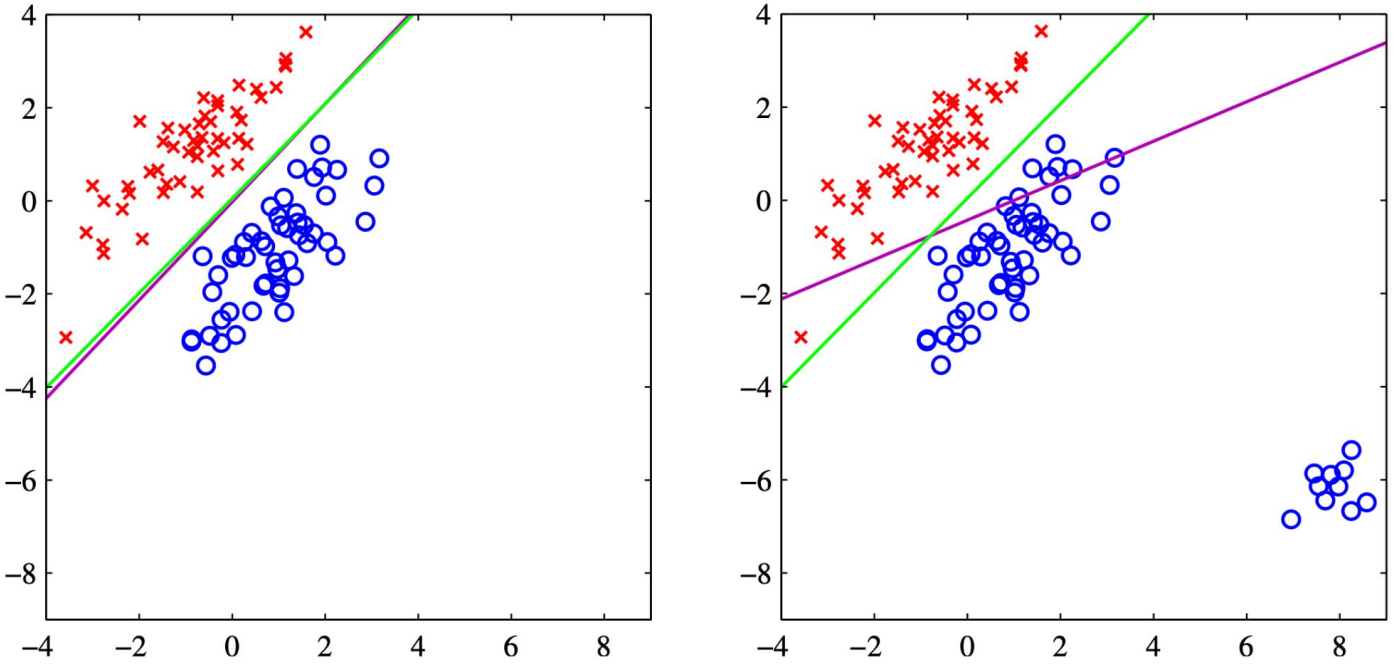
Least Squares for Classification

The error function can be conveniently written as

$$E_D(\widetilde{\mathbf{W}}) = \frac{1}{2}\mathrm{Tr}\left\{(\widetilde{\mathbf{X}}\widetilde{\mathbf{W}} - \mathbf{T})^{\mathrm{T}}(\widetilde{\mathbf{X}}\widetilde{\mathbf{W}} - \mathbf{T})\right\}$$

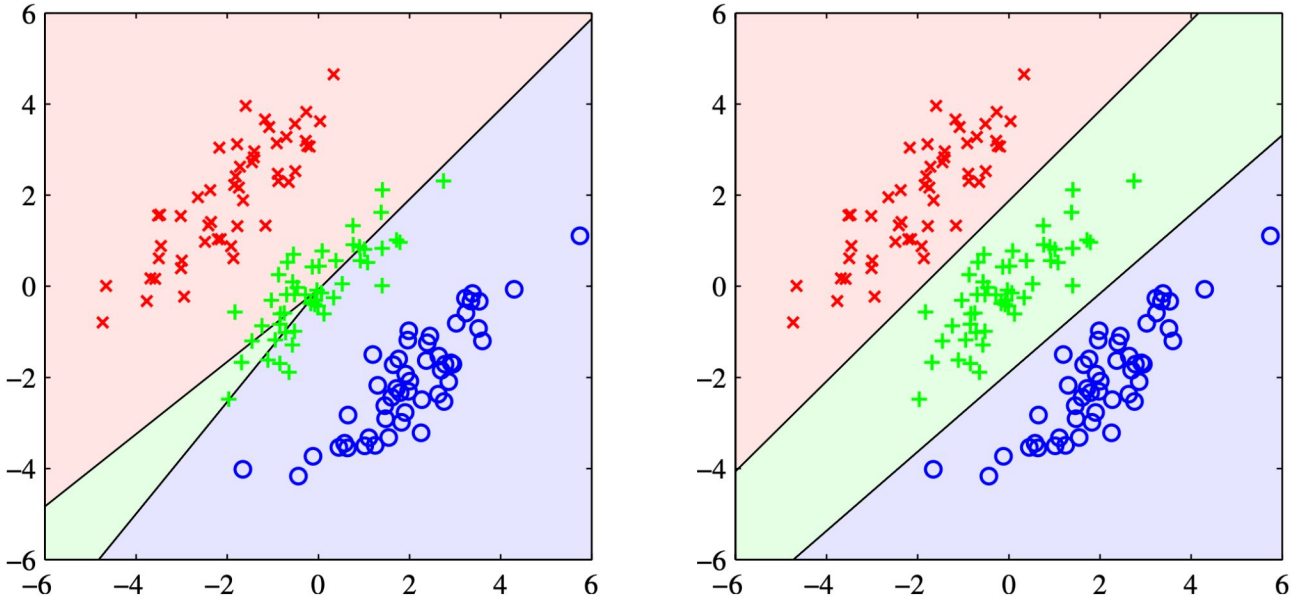
Minimize $E_D(\widetilde{\mathbf{W}})$ as a function of $\widetilde{\mathbf{w}}$:

Least Squares Issues - Outliers



Magenta - LS classifier
Green - Logistic Regression classifier

Least Squares Issues - Masking



Left - LS classifier
Right - Logistic Regression classifier

Least Squares Issues - Predictions ≠ Probabilities

$\mathbf{y}_{LS}(\mathbf{x})$ are not probabilities

Rough

Next
The Perceptron