CSE 575: Statistical Machine Learning (Spring 2021)
Instructor: Nupur Thakur

Supervised Learning

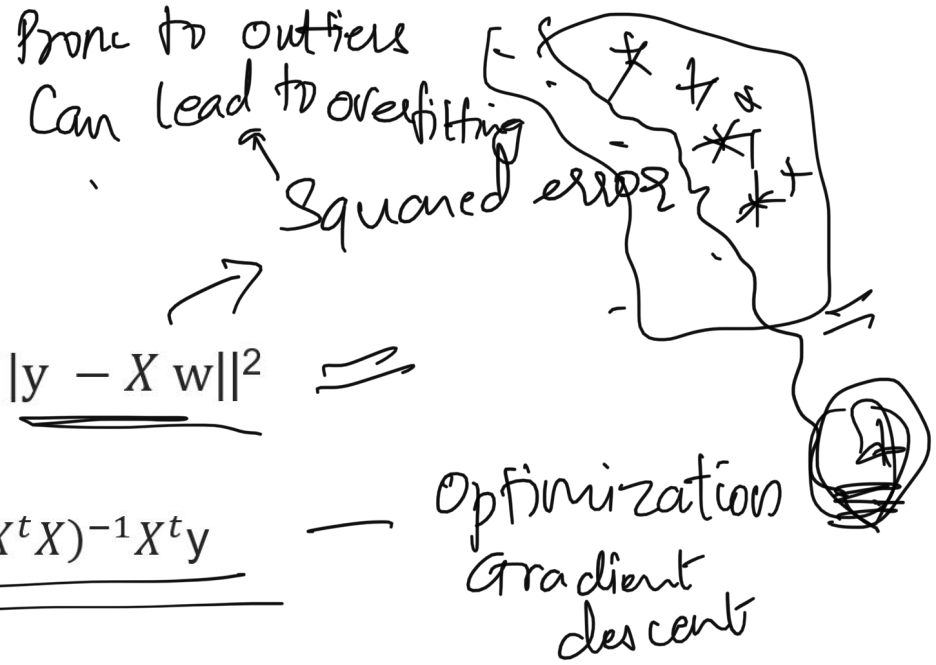
Table of contents

Linear Regression

- Regression - A training set of n samples $<x^{(i)}, y^{(i)}>$ where $y^{(i)}$ is a continuous “label” (or target value) for $x^{(i)}$
- Linear regression - modeling the relation between y and x via a linear function

$$y \approx w_0 + w_1 x_1 + \ldots + w_d x_d = \mathbf{w}^t \mathbf{x}$$

Linear Regression



- The error is given as -

$$\text{argmin} \, ||e||^2 = ||y - Xw||^2$$

- Weights can be found using -

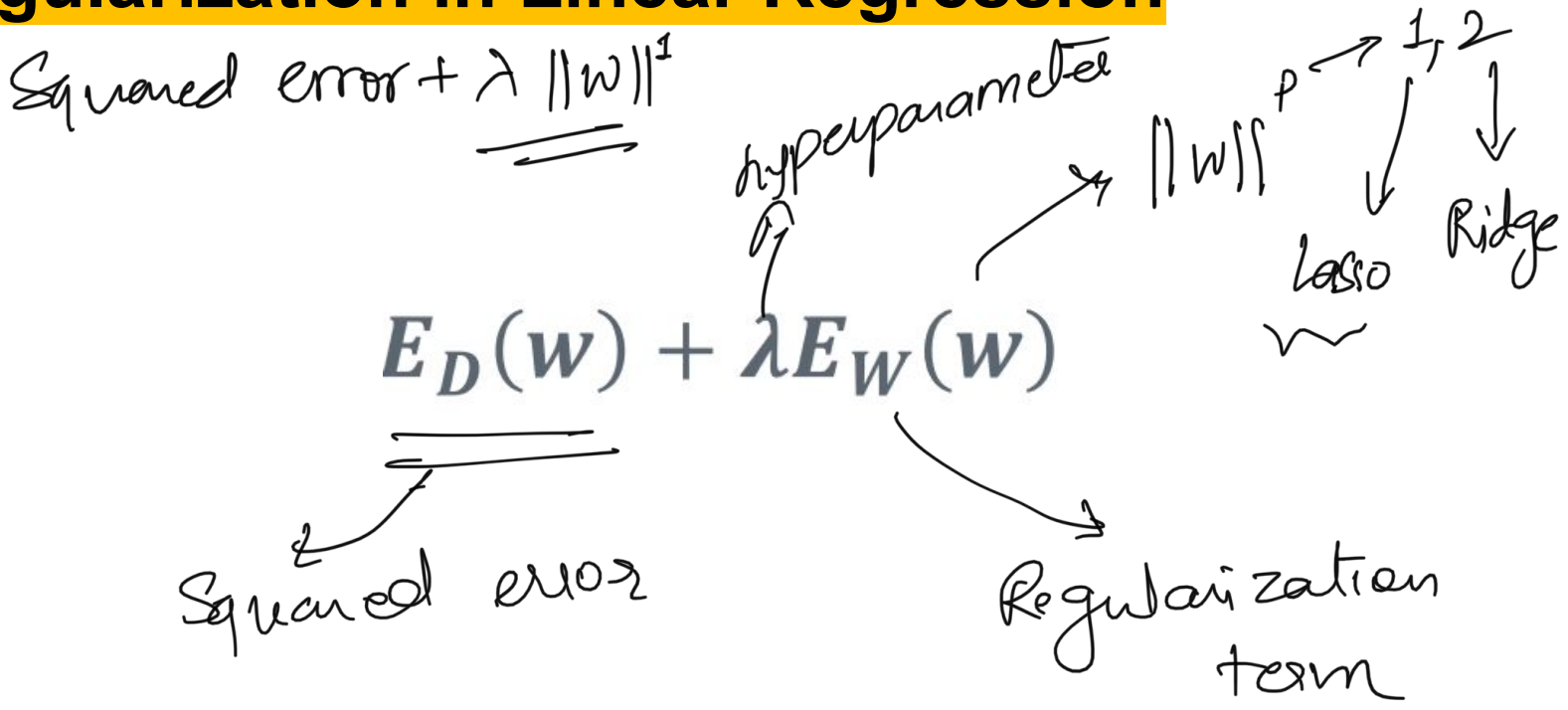
$$\hat{w} = (X^t X)^{-1} X^t y$$

Optimization
Gradient descent

- How to generalize?

$$y = w_0 + w_1 \phi_1(\mathbf{x}) + \ldots + w_{M-1} \phi_{M-1}(\mathbf{x})$$

Regularization in Linear Regression



Why regularize and how it works?

- Used to avoid overfitting of the network

Covariance M

[ ]

- Regularization terms shrinks the w estimates.

100 X 2

	Mean	SD
100		

Test Accuracy	Logistic	Naive
0	. %	%
1		

Questions?