Finance Data Science
Lecture 10: Machine Learning Review

Laurent El Ghaoui

MFE 230P, Summer 2017
MFE Program
Haas School of Business
UC Berkeley

7/10/2017

Outline

Machine Learning Review

Topics:

- Data points and features
- Unsupervised vs. supervised learning
- Performance metrics
- Loss functions
- Regularization
- Kernels
- Neural networks

Neural networks

Adding one layer to least-squares

Regularized LS:

$$\min_w \ \|X^T w - y\|_2^2 + \rho^2 \|w\|_2^2$$

with

- $X \in \mathbf{R}^{n \times m}$ the data matrix (each column is a point);
- $y \in \mathbf{R}^m$ is the response vector;
- w weights the different features;
- Prediction rule: $\hat{y}(x) = w^T x$.

The above is convex.

Adding one layer

We now add a layer:

$$\min_{w,W,Z} \|Z^T w - y\|_2^2 + \rho^2 \|w\|_2^2 + \theta^2 \|W\|_F^2 \ : \ Z = \phi(W^T X)$$

- new variables: Z is a proxy data matrix, W is a weighting matrix;
- ϕ is the RELU function: $\phi(V) = max(V, 0)$, acting component-wise on a matrix V;
- $\theta > 0$ is a regularization parameter.
- Prediction rule: $\hat{y}(x) = w^T z$, where $z = \phi(W^T x)$.

- New problem is not convex;
- Solved by eliminating Z and minimizing over weights W, w.