NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas
Department : E & ECE, IIT Kharagpur

Topic

Lecture 17: Optimization in ML

CONCEPTS COVERED

Concepts Covered:

- Optimization
  - Stochastic Gradient Descent
  - Batch Optimization
  - Mini-batch optimization
- Optimization in ML
- Linear and Logistic Regression
- Softmax classifier
- Nonlinearity

Optimization in Machine Learning

Optimization in Machine Learning

- Goal of optimization is to reduce a cost function *J(W)* to optimize some performance measure *P.*
- In pure optimization minimizing *J* is the goal in and of itself.
- In Machine Learning *J(W)* is minimized w.r.t parameter *W* on training data (training error), and we the error to be low on unforeseen (test) data.
- Test error (generalization error) should be low.

Optimization in Machine Learning

Assumptions

- Test and Training data are generated by a probability distribution: Data generating process.
- Data samples in each data set are independent.
- Training set and Test set are identically distributed.

Performance of ML is its ability to

- Make the training error small.
- Reduce the gap between training and test error.

Underfitting and Overfitting

- **Underfitting**: Model is not able to obtain sufficiently low training error.
- **Overfitting**: The gap between training and test error is too large.

We can control Overfitting/ Underfitting by altering its Capacity

Set of functions the learning algorithm can select as being the solution

Linear and Logistic Regression

Linear & Logistic Regression- Binary Classification

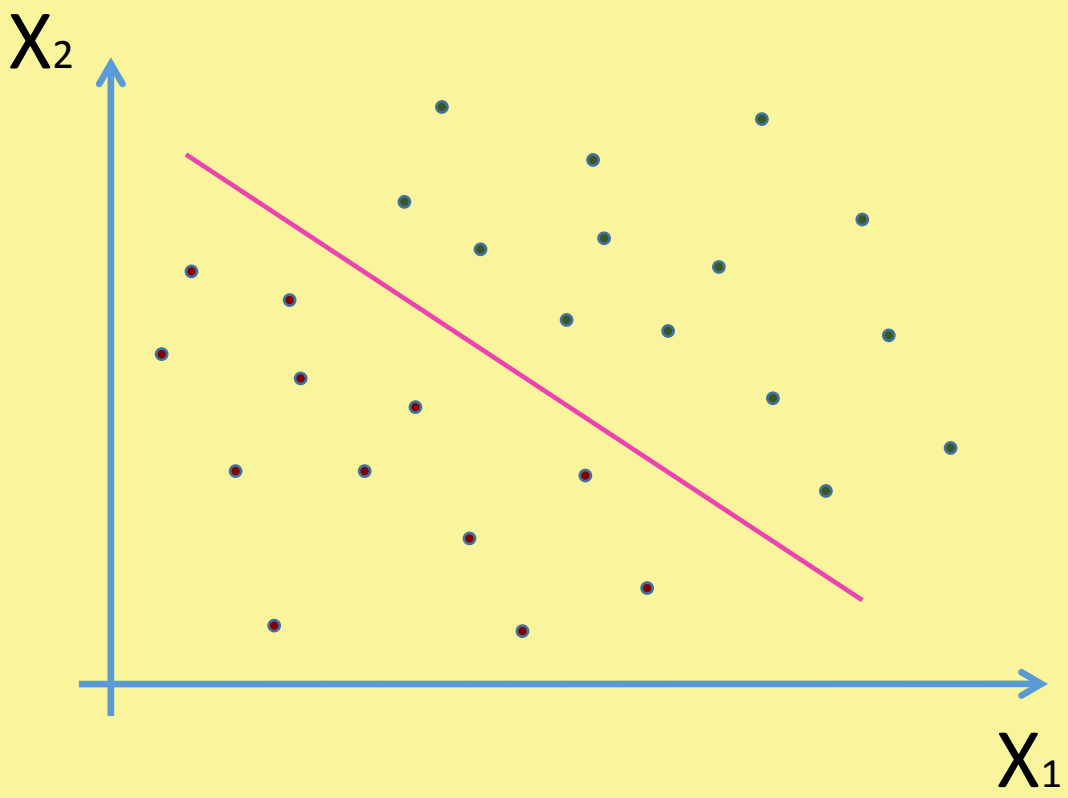
Linear Regression

$$f : X \in R^d \rightarrow y \in R \qquad \hat{y} = W^t X$$

Logistic Regression

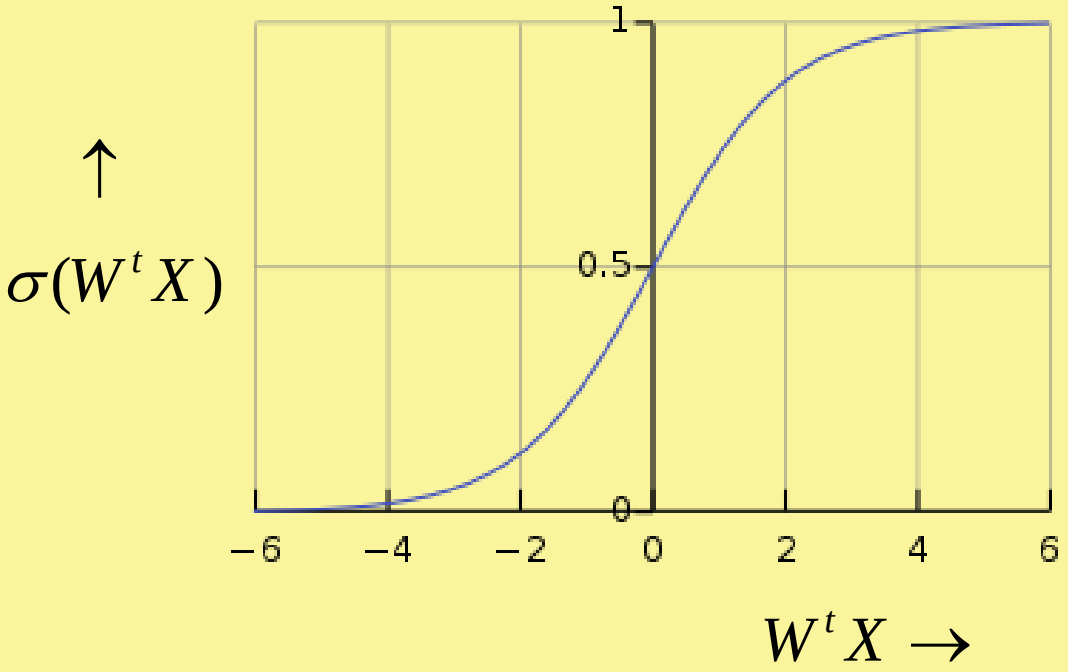
$$p(y \mid X; W) = \sigma(W^t X)$$

Linear Regression



Logistic Regression

$$\sigma(W^t X) = \frac{1}{1+e^{-W^t X}} \Rightarrow$$



Softmax Classifier

- Generalization of Binary Logistic Classifier to Multiple Classes

$$s_{y_i} = f(X_i, W)_{y_i} = (WX_i)_{y_i} = W_{y_i}^t X_i$$

- Softmax Classifier

$$p(y_i \mid X_i; W) = \frac{e^{s_{y_i}}}{\sum_{\forall j} e^{s_j}}$$

NPTEL ONLINE CERTIFICATION COURSES

Thank you