

# Lecture notes on Machine learning

## Contents

### 1 Lecture 1 : Introduction to Machine Learning

### 2 Lecture 2

- 2.1 Solving Least Squares in General (for Linear models)

### 3 Lecture 3 : Regression

- 3.1 Linear regression
- 3.2 Least square solution
- 3.3 Geometrical interpretation of least squares

### 4 Lecture 4 : Least Squares Linear Regression

- 4.1 Least Square Linear Regression Model
- 4.2 Level Curves and Surfaces
- 4.3 Gradient Vector
- 4.4 Directional Derivative
- 4.5 Hyperplane
- 4.6 Tangential Hyperplane
- 4.7 Gradient Descent Algorithm
- 4.8 Local Minimum and Local Maximum

### 5 Lecture 5 : Convex functions

- 5.1 Recap
- 5.2 Point 1
- 5.3 Point 2
- 5.4 Point 3
- 5.5 Point 4
  - 5.5.1 Overfitting
  - 5.5.2 Next problem
- 5.6 Point 5

### 6 Lecture 6 : Regularized Solution to Regression Problem

- 6.1 Problem formulation
- 6.2 Duality and KKT conditions
- 6.3 Bound on  $\lambda$  in the regularized least square solution

- 6.4 RMS Error variation
- 6.5 Alternative objective function
  - A review of probability theory
  - 6.5.1 The three axioms of probability
  - 6.5.2 Bayes' theorem
  - 6.5.3 Independent events

## **7 Lecture 7 : Probability**

- 7.1 Note
- 7.2 Part of speech(pos) example
- 7.3 Probability mass function(pmf) and probability density function(pdf)
  - 7.3.1 Joint distribution function
  - 7.3.2 Marginalization
- 7.4 Example
- 7.5 Conditional Density
- 7.6 Expectation
  - 7.6.1 Properties of  $E(x)$
- 7.7 Variance
- 7.8 Covariance
  - 7.8.1 Properties of Covariance
- 7.9 Chebyshev's Inequality
- 7.10 Bernoulli Random Variable
- 7.11 Binomial Random Variable
- 7.12 Central Limit Theorem
- 7.13 Maximum Likelihood and Estimator
- 7.14 Bayesian estimator

## **8 Lecture 8**

- 8.1 Bernoulli Distribution
- 8.2 Bayesian Estimation

## **9 Lecture 9 : Multinomial Distribution**

- 9.0.1 Posterior probability
- 9.0.2 Summary
- 9.1 Gaussian Distribution
  - 9.1.1 Information Theory
  - 9.1.2 Expectation for  $I(X=x)$ :
  - 9.1.3 Observations:
  - 9.1.4 Properties of gaussian univariate distribution

## **10 Lecture 10 : Multivariate Gaussian Distribution**

- 10.1 Multivariate Gaussian Variable
  - 10.1.1 Unbiased Estimator
- 10.2 Dealing with Conjugate Priors for Multivariate Gaussian

**11 Lecture 11**

- 11.1 Recall
- 11.2 Bayes Linear Regression
- 11.3 Pure Bayesian - Regression
- 11.4 Sufficient Statistic
- 11.5 Lasso

**12 Lecture 12 : Bias-Variance tradeoff**

- 12.1 Expected Loss

**13 Lecture 13**

- 13.1 Conclude Bias-Variance
  - 13.1.1 Summary
  - 13.1.2 Bayesian Linear Regression(BLR)
  - 13.1.3 General Problems with Standard Distribution
- 13.2 Empirical Bayes
  - 13.2.1 First Approach: Approximate the posterior
  - 13.2.2 Second Approach: Empirical Bayes
  - 13.2.3 Solve the eigenvalue equation

**14 Lecture 14 : Introduction to Classification****15 Lecture 15: Linear Models for Classification**

- 15.1 Generalized linear models
- 15.2 Three broad types of classifiers
  - 15.2.1 Examples
- 15.3 Handling Multiclass
  - 15.3.1 Avoiding ambiguities
- 15.4 Least Squares approach for classification
  - 15.4.1 Limitations of Least Squares

**16 Lecture 16]**

- 16.1 Introduction
- 16.2 Problems of linear regression
  - 16.2.1 Sensitivity to outliers
  - 16.2.2 Masking
- 16.3 Possible solutions
- 16.4 Summary

**17 Lecture 17****18 Lecture 18: Perceptron**

- 18.1 Fisher's discriminant
- 18.2 Perceptron training
  - 18.2.1 Intuition

**19 Lecture 19**

- 19.1 Introduction
- 19.2 Margin
- 19.3 Support Vector Machines
- 19.4 Support Vectors
- 19.5 Objective Design in SVM
  - 19.5.1 Step 1: Perfect Separability
  - 19.5.2 Step 2: Optimal Separating Hyperplane For Perfectly Separable Data
  - 19.5.3 Step 2: Separating Hyperplane For Overlapping Data

**20 Lecture 20: Support Vector Machines (SVM)**

- 20.1 Recap
- 20.2 Distance between the points
- 20.3 Formulation of the optimization problem
- 20.4 Soft Margin SVM
  - 20.4.1 Three types of  $g$  points
- 20.5 Primal and Dual Formulation
  - 20.5.1 Primal formulation
  - 20.5.2 Dual Formulation
- 20.6 Duality theory applied to KKT

**21 Lecture 21: The SVM dual**

- 21.1 SVM dual
- 21.2 Kernel Matrix
  - 21.2.1 Generation of  $\phi$  space
- 21.3 Requirements of Kernel
  - 21.3.1 Examples of Kernels
- 21.4 Properties of Kernel Functions

**22 Lecture 22: SVR and Optimization Techniques**

- 22.1 Other occurrence of kernel
  - 22.1.1 Some variants of SVM's
- 22.2 Support Vector Regression
- 22.3  $L_1$  SVM
- 22.4 Kernel Adatron

**23 Lecture 23**

- 23.1 Sequential minimization algorithm - SMO
- 23.2 Probabilistic models

**24 Lecture 24: Prob. Classifiers**

- 24.1 Non Parametric Density Estimation
- 24.2 Parametric Density Estimation

**25 Lecture 25**

- 25.1 Exponential Family Distribution**
- 25.2 Discrete Feature Space**
- 25.3 Naive Bayes Assumption**
- 25.4 Graphical Models**
- 25.5 Graphical Representation of Naive Bayes**
- 25.6 Graph Factorisation**
- 25.7 Naive Bayes Text Classification**