Machine Learning

1 Introduction

Machine Learning is the science about learning from past data and then making prediction about the future, which is driven by:

- Statistical modelling
- Data analysis
- Numerical optimization

2 Machine Learning Algorithms Family

- Supervised Learning: the learning process is led by the labels associated to the features
  - Classification: the outputs are discrete values (binary or multi-class) and classified in labels
    - Support Vector Machine (SVM)
    - Discriminant Analysis
    - Naive Bayes
    - Nearest Neighbor
  - Regression: the outputs are continuous values, while n measured quantities correlate all the variables
    - Linear Regression, GLM
    - SVR, GPR
    - Ensemble Methods
    - Decision Trees
    - Neural Networks
- Unsupervised Learning: there are no outputs available in learning process, then the the model is build by recognizing common patterns and label them in clusters (clustering)
  - K-Means, K-Medoids, Fuzzy logic, C-Means

  - Hierarchical
  - Gaussian Mixture
  - Neural Networks
  - Hidden Markov Model

- Reinforcement Learning
  -

3 Regression Algorithms

3.1 Linear Regression

The linear regression is linear in parameters even if can be nonlinear in input-output relation. The general model is a linear combination of the model parameters $\theta = [\theta_1 \dots \theta_n]^T$ in x:

$$y = \sum_{i=1}^{n} \theta_i x_i = \theta^T x \tag{1}$$

where: n = dimension of the model (number of features)

3.2 Linear Regression

The linear regression is linear in parameters even if can be nonlinear in input-output relation. The general model is a linear combination of the model parameters $\theta = [\theta_1 \dots \theta_d]$ in $\phi(x)$ in $\phi(x)$:

$$y = \sum_{j=1}^{d} \theta_j \phi_j(x) + \epsilon = \theta^T \phi(x) + \epsilon \tag{2}$$

where: d = order of the model, $\phi(x) = [\phi_1(x) \dots \phi_d(x)]$ vector-valued map which columns vectors are the basis functions, $\epsilon \sim N(0, \sigma^2)$ stochastic error.

4 Classification Algorithms

4.1 Naive Bayes Classifier

4.2 Support Vector Machine

5 Validation

5.1 Bayesan Loss

For a given estimate of the parameters, the cost/loss function measures the error in the prediction:

- Quadratic loss: $L = (\theta - \hat{\theta})^2$
- Absolute-value loss: $L = |\theta - \hat{\theta}|$
- Hit-or-miss loss: $L = \begin{cases} 0, & |\theta - \hat{\theta}| \ll \delta \\ 1, & |\theta - \hat{\theta}| > \delta \end{cases}$
- Huber loss: $L = \begin{cases} 0, & (\theta - \hat{\theta}) \ll \delta \\ 1, & (\theta - \hat{\theta}) > \delta \end{cases}$

5.2 Bayesan Loss Estimator

- Minimum Mean-Square Error (MMSE)
- Minimum Absolute Error (MAE)
- Maximum A-Posteriori Estimator (MAP)
- Maximum Likelihood Estimator (ML)