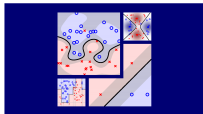


# Machine Learning Techniques (機器學習技法)



Lecture 1126: Summary

Hsuan-Tien Lin (林軒田)

htlin@csie.ntu.edu.tw

Department of Computer Science  
& Information Engineering

National Taiwan University  
(國立台灣大學資訊工程系)



# Agenda

Lecture 1126: Summary

# Model 1: PLA

- err: 0/1
- $\widehat{\text{err}}$ :  $\max(-ys, 0)$
- optimization:
  - SGD on  $\widehat{\text{err}}$
  - special minimizer of err when linear separable
- $\Phi$ : identity
- regularization/validation: none
- parameters: none
- practical use: online learning, and ‘teaching’

## Model 2: pocket

- err: 0/1
- $\widehat{\text{err}}$ : none
- optimization: special random search with PLA as core
- $\Phi$ : identity
- regularization/validation: none
- parameters: number of iterations  $T$
- practical use: seldom

## Model 3: ridge linear regression

- err: squared
- $\widehat{\text{err}}$ : squared
- optimization: analytic solution
- $\Phi$ : identity
- regularization/validation: L2 regularization
- parameters: regularization level  $\lambda$
- practical use: for 'decent' baseline solution

## Model 4: logistic regression

- $\text{err}$ : cross entropy
- $\widehat{\text{err}}$ : cross entropy
- optimization: GD/SGD (basic)
- $\Phi$ : identity
- regularization/validation: often L2 regularization, or L1 if needing sparsity
- parameters: regularization level  $\lambda$  and GD/SGD iterations  $T$ , step size  $\eta$
- practical use: very useful for baseline hard/soft classification

## Model 5: ridge polynomial regression

- err: squared
- $\widehat{\text{err}}$ : squared
- optimization: analytic solution
- $\Phi$ : polynomial
- regularization/validation: L2 regularization
- parameters: regularization level  $\lambda$  and polynomial degree  $Q$
- practical use: often for (1-D) regression problems

## Model 6: soft-margin SVM

- err: 0/1
- $\widehat{\text{err}}$ : hinge
- optimization: (special) quadratic programming
- $\Phi$ : embedded in kernel  $K$
- regularization/validation: large-margin/leave-one-out bound or CV
- parameters: error penalty rate  $C$  and kernel parameters
- practical use: very popular for classification



## Model 7: kernel logistic regression

- $\text{err}$ : cross entropy
- $\widehat{\text{err}}$ : cross entropy
- optimization: GD/SGD
- $\Phi$ : embedded in kernel  $K$
- regularization/validation: L2 regularization
- parameters: regularization level  $\lambda$  and kernel parameters
- practical use: Probabilistic SVM more often used

## Model 8: kernel ridge regression

- err: squared
- $\widehat{\text{err}}$ : squared
- optimization: analytic solution (but often solved iteratively)
- $\Phi$ : embedded in kernel  $K$
- regularization/validation: L2 regularization
- parameters: regularization level  $\lambda$  and kernel parameters
- practical use: only on small data

# Model 9: AdaBoost

- $\text{err}$ : 0/1
- $\widehat{\text{err}}$ : exponential
- optimization: functional gradient descent with the help of base algorithm
- $\Phi$ : diverse hypotheses found iteratively
- regularization/validation: often through early stopping
- parameters: number of iterations  $T$
- practical use: 'boost' decision trees/stumps

## Model 10: Decision Tree

- err: 0/1 or squared
- $\widehat{\text{err}}$ : not fully clear
- optimization: heuristic greedy
- $\Phi$ : conditional hypotheses found recursively
- regularization/validation: pruning
- parameters: lots of heuristic choices
- practical use: 'explainable' nonlinear model

# Model 11: Bagging/Random Forest

- err: squared, or 'any'?!
  - $\widehat{\text{err}}$ : squared
- optimization: through base algorithm/decision trees
- $\Phi$ : diverse hypotheses through bootstrapping, and random projection/combination
- regularization/validation: variance decreasing/OOB error
- parameters: number of iterations  $T$
- practical use: 'stabelize' any model/trees

## Model 12: Gradient Boosted Decision Tree

- $\text{err}$ : squared, or any
- $\widehat{\text{err}}$ : squared, or any
- optimization: functional gradient descent with the help of base algorithm
- $\Phi$ : diverse hypotheses found iteratively
- regularization/validation: often through early stopping
- parameters: number of iterations  $T$
- practical use: very popular for information retrieval

## Model 13: Neural Networks/Deep Learning

- $err$ : squared, or cross entropy
- $\widehat{err}$ : squared, or cross entropy
- optimization: GD/SGD with help of backprop, initializing with auto-encoder-like pre-training
- $\Phi$ : learned and represented by hidden neurons
- regularization/validation: weight elimination, early stopping
- parameters: number of iterations  $T$ , step size  $\eta$ , network architecture
- practical use: very popular nowadays for vision/speech

## Model 14: RBF Networks

- err: squared, or 0/1
- $\widehat{\text{err}}$ : squared
- optimization: prototype-finding with  $k$ -means-like algorithms, then analytic solution
- $\Phi$ : learned and represented by prototypes
- regularization/validation: number of prototypes
- parameters: number of prototypes  $k$ , RBF width  $\gamma$
- practical use: less popular nowadays



# Summary

## Lecture 1126: Summary