CS 340: Review

AD

April 2011

CS340

- K-Nearest Neighbours

CS340

- K-Nearest Neighbours
- Principal Components Analysis

- K-Nearest Neighbours
- Principal Components Analysis
- Markov Chains and PageRank

CS340

- K-Nearest Neighbours
- Principal Components Analysis
- Markov Chains and PageRank
- Generative/Bayesian Classifiers

- K-Nearest Neighbours
- Principal Components Analysis
- Markov Chains and PageRank
- Generative/Bayesian Classifiers
- Linear Regression

- K-Nearest Neighbours
- Principal Components Analysis
- Markov Chains and PageRank
- Generative/Bayesian Classifiers
- Linear Regression
- Discriminative/Logistic Regression

CS340

- K-Nearest Neighbours
- Principal Components Analysis
- Markov Chains and PageRank
- Generative/Bayesian Classifiers
- Linear Regression
- Discriminative/Logistic Regression
- Neural Networks

- K-Nearest Neighbours
- Principal Components Analysis
- Markov Chains and PageRank
- Generative/Bayesian Classifiers
- Linear Regression
- Discriminative/Logistic Regression
- Neural Networks
- K-Means

CS340

- K-Nearest Neighbours
- Principal Components Analysis
- Markov Chains and PageRank
- Generative/Bayesian Classifiers
- Linear Regression
- Discriminative/Logistic Regression
- Neural Networks
- K-Means
- Mixture Models and HMM

At this stage, you should...

- Know what class of methods to use when given a problem.

At this stage, you should...

- Know what class of methods to use when given a problem.
- Know the criteria optimized by these methods.

At this stage, you should...

- Know what class of methods to use when given a problem.
- Know the criteria optimized by these methods.
- Be able to do basic linear algebra and proba/stats calculations.

At this stage, you should...

- Know what class of methods to use when given a problem.
- Know the criteria optimized by these methods.
- Be able to do basic linear algebra and proba/stats calculations.
- Be able to provide/complete simple Matlab scripts.

K-Nearest Neighbours

- Supervised classification.

K-Nearest Neighbours

- Supervised classification.
- Metric / K.

K-Nearest Neighbours

- Supervised classification.
- Metric / K.
- Strengths: easy to implement/understand, can perform reasonably well on clean data.

K-Nearest Neighbours

- Supervised classification.
- Metric / K.
- Strengths: easy to implement/understand, can perform reasonably well on clean data.
- Weakness: curse of dimensionality in high dimension, basic version non-probabilistic, cannot handle missing features.

K-Nearest Neighbours

- Supervised classification.
- Metric / K.
- Strengths: easy to implement/understand, can perform reasonably well on clean data.
- Weakness: curse of dimensionality in high dimension, basic version non-probabilistic, cannot handle missing features.
- How to select the metric? how to select K?

PCA

- Dimensionality reduction.

PCA

- Dimensionality reduction.
- Minimize

$$\sum_{i=1}^{N} \left\| \mathbf{x}_i - \widehat{\mathbf{x}}_i \right\|^2$$

PCA

- Dimensionality reduction.
- Minimize
$$\sum_{i=1}^{N} \left\| \mathbf{x}_i - \widehat{\mathbf{x}}_i \right\|^2$$
- We obtain
$$\widehat{\mathbf{x}}_i = \sum_{j=1}^{k} \left(\mathbf{x}_i^{\mathrm{T}} \ \mathbf{u}_i \right) \ \mathbf{u}_i$$

PCA

- Dimensionality reduction.
- Minimize

$$\sum_{i=1}^{N} \left\| \mathbf{x}_i - \widehat{\mathbf{x}}_i \right\|^2$$

- We obtain

$$\widehat{\mathbf{x}}_i = \sum_{j=1}^{k} \left(\mathbf{x}_i^{\mathsf{T}} \ \mathbf{u}_i \right) \ \mathbf{u}_i$$

- Can be used for latent semantic analysis, visualization, compression.

PCA

- Dimensionality reduction.
- Minimize

$$\sum_{i=1}^{N} \left\| \mathbf{x}_i - \widehat{\mathbf{x}}_i \right\|^2$$

- We obtain

$$\widehat{\mathbf{x}}_i = \sum_{j=1}^{k} \left(\mathbf{x}_i^{\mathsf{T}} \ \mathbf{u}_i \right) \ \mathbf{u}_i$$

- Can be used for latent semantic analysis, visualization, compression.
- Usefulness in nonlinear scenarios? How to select k?

- What is a Markov chain?

- What is a Markov chain?
- When do we have $\lim_k \pi_k = \pi$? when is this limit unique?

Markov Chains and PageRank

- What is a Markov chain?
- When do we have $\lim_k \pi_k = \pi$? when is this limit unique?
- What is the principle of PageRank?

Markov Chains and PageRank

- What is a Markov chain?
- When do we have $\lim_k \boldsymbol{\pi}_k = \boldsymbol{\pi}$? when is this limit unique?
- What is the principle of PageRank?
- Power method.

- What is a Markov chain?

- What is a Markov chain?
- When do we have $\lim_k \boldsymbol{\pi}_k = \boldsymbol{\pi}$? when is this limit unique?

Markov Chains and PageRank

- What is a Markov chain?
- When do we have $\lim_k \boldsymbol{\pi}_k = \boldsymbol{\pi}$? when is this limit unique?
- What is the principle of PageRank?

Markov Chains and PageRank

- What is a Markov chain?
- When do we have $\lim_k \pi_k = \pi$? when is this limit unique?
- What is the principle of PageRank?
- Power method.

Bayesian Classifiers

- Supervised classification.

Bayesian Classifiers

- Supervised classification.
- Require modelling class conditional densities of features and use Bayes rules.

Bayesian Classifiers

- Supervised classification.
- Require modelling class conditional densities of features and use Bayes rules.
- Learning of the parameters using MLE, MAP (or full Bayes): multinomial, Gaussian.

Bayesian Classifiers

- Supervised classification.
- Require modelling class conditional densities of features and use Bayes rules.
- Learning of the parameters using MLE, MAP (or full Bayes): multinomial, Gaussian.
- In multivariate Gaussian case, same structure as logistic but different parameters.

Bayesian Classifiers

- Supervised classification.
- Require modelling class conditional densities of features and use Bayes rules.
- Learning of the parameters using MLE, MAP (or full Bayes): multinomial, Gaussian.
- In multivariate Gaussian case, same structure as logistic but different parameters.
- Strengths: Exploit features density, missing features easily handled.

Bayesian Classifiers

- Supervised classification.
- Require modelling class conditional densities of features and use Bayes rules.
- Learning of the parameters using MLE, MAP (or full Bayes): multinomial, Gaussian.
- In multivariate Gaussian case, same structure as logistic but different parameters.
- Strengths: Exploit features density, missing features easily handled.
- Weakness: Require reliable model of class conditional density, can be difficult to learn.

Linear Regression

- Least square criterion, regression and probabilistic interpretation.

Linear Regression

- Least square criterion, regression and probabilistic interpretation.
- Use of basis functions.

Linear Regression

- Least square criterion, regression and probabilistic interpretation.
- Use of basis functions.
- Limitations of least square regression: robust regression, ridge regression, L1 regression

Linear Regression

- Least square criterion, regression and probabilistic interpretation.
- Use of basis functions.
- Limitations of least square regression: robust regression, ridge regression, L1 regression
- How to estimate the regularization coefficient?

Logistic Regression

- Discriminative model for supervised classification.

Logistic Regression

- Discriminative model for supervised classification.
- Use of basis functions.

Logistic Regression

- Discriminative model for supervised classification.
- Use of basis functions.
- Maximum Likelihood/Penalized Maximum likelihood estimation via gradient.

Logistic Regression

- Discriminative model for supervised classification.
- Use of basis functions.
- Maximum Likelihood/Penalized Maximum likelihood estimation via gradient.
- Geometry of decision boundaries.

Logistic Regression

- Discriminative model for supervised classification.
- Use of basis functions.
- Maximum Likelihood/Penalized Maximum likelihood estimation via gradient.
- Geometry of decision boundaries.
- Strengths: Quite easy to implement and probabilistic model.

Logistic Regression

- Discriminative model for supervised classification.
- Use of basis functions.
- Maximum Likelihood/Penalized Maximum likelihood estimation via gradient.
- Geometry of decision boundaries.
- Strengths: Quite easy to implement and probabilistic model.
- Weakness: Does not exploit density of features, cannot handle missing features.

K-Means

- Unsupervised learning: clustering.

K-Means

- Unsupervised learning: clustering.
- Objective function minimized.

K-Means

- Unsupervised learning: clustering.
- Objective function minimized.
- Strengths: Fast algorithm

K-Means

- Unsupervised learning: clustering.
- Objective function minimized.
- Strengths: Fast algorithm
- Weaknesses: non probabilistic, local maxima, cluster shapes not modelled...

K-Means

- Unsupervised learning: clustering.
- Objective function minimized.
- Strengths: Fast algorithm
- Weaknesses: non probabilistic, local maxima, cluster shapes not modelled...
- How to estimate the number of clusters?

Mixture Models and EM

- Unsupervised learning: clustering and density estimation.

Mixture Models and EM

- Unsupervised learning: clustering and density estimation.
- Maximum likelihood parameter learning using EM.

Mixture Models and EM

- Unsupervised learning: clustering and density estimation.
- Maximum likelihood parameter learning using EM.
- EM exploits that complete log-likelihood can be easily maximized.

Mixture Models and EM

- Unsupervised learning: clustering and density estimation.
- Maximum likelihood parameter learning using EM.
- EM exploits that complete log-likelihood can be easily maximized.
- Strengths: Flexible models, elegant algorithm.

Mixture Models and EM

- Unsupervised learning: clustering and density estimation.
- Maximum likelihood parameter learning using EM.
- EM exploits that complete log-likelihood can be easily maximized.
- Strengths: Flexible models, elegant algorithm.
- Weaknesses: local maxima...