CSCI 181 / E-181 Spring 2014

2nd midterm review

David Wihl
davidwihl@gmail.com

April 26, 2014

1 Support Vector Machines

1.1 Max-Margin Classification

SVMs are based on three "big ideas":

- *margin.* Maximizes distance between the closest points
- *duality.* Take a hard problem and transform it into an easier problem to solve.
- *kernel trick.* Map input vectors to higher dimensional, more expressive features.

Characteristics of SVMs:

- *linearly separable.* assumes that linear separation is possible
- *convex optimization.* SVM originated as a backlash against neural nets due to non-convexity. In Neural Nets, results were often non-reproducible as different researchers found different results.
- *global optimum* SVMs will find the global optimum.

DATA: $\{x_n, t_n\}_{n=1}^{N}, t_n \in \{-1, +1\}$

J Basis functions: $\phi_j(x) \rightarrow \Re$

Vector function: $\Phi X \rightarrow \Re^J$ (column vector)

Assume linear separability

Objective function: $f(\vec{x}, \vec{w}, b) = \phi(\vec{x})^\intercal \vec{w} + b$ where b is the bias or offset. The sign of $f(\dot{})$ will determine classification $(-1, +1)$

SVM is used as a classifier, such that:

$$y(\vec{x}, \vec{w}, b) \quad = \quad \begin{cases} +1, & \text{if } \vec{\Phi}(\vec{x})^\intercal \vec{w} + b > 0 \\ -1, & \text{otherwise} \end{cases}$$

Decision boundary is the hyperplane where $\vec{\phi}(\vec{x})^\intercal \vec{w} + b = 0$

Sources:

1. Lecture 14
2. Bishop 6.0-6.2
3. Course notes - maxmargin
4. Section 7 review

1.2 SVM

Lecture 15

Bishop 7.0-7.1

Section 8

2 Markov Decision Processes

Lecture 16

Course notes - MDP

Section 9

2.1 Partially Observable MDP

Course notes - POMDP

Section 10

2.2 Hidden Markov Models

Bishop 13.0-13.2

2.3 Mixture Models

Bishop 9.0-9.2

3 Reinforcement Learning

Course notes - RL

Section 9

3.1 Value and Policy Iteration

Lecture 17

Course notes - policyiter

4 Expectation Maximization

Bishop 9.3

Section 11