Towards Bayesian Black Box Learning Systems

Jo-Anne Ting, USC

Women in Machine Learning Workshop
October 4, 2006
Collaborators

- Aaron D’Souza
- Stefan Schaal

Robotics
- Jun Nakanishi
- Michael Mistry
- Jan Peters

Neuroscience
- Donna Hoffman
- Shinji Kakei
- John Kalaska
- Mitsuo Kawato

- Lauren Sergio
- Peter Strick
- Kenji Yamamoto
- Toshinori Yoshioka
Our Research

• Motivation:
  • Why create Bayesian black box systems?

• Past work:
  • The simplest scenario: linear regression
  • A more realistic scenario: noisy input data

• Our latest work:
  • Moving to nonlinear functions
Why create Bayesian Black Box Systems?

- Black box systems allow autonomous learning
  - Useful in vehicular, surveillance, biological & robotic systems etc.

- Bayesian techniques allow elimination of open parameters
  - No need for tuning!
Our Research

• Motivation:
  • Why create Bayesian black box systems?
• Past Work:
  • The simplest scenario: linear regression
  • A more realistic scenario: noisy input data
• Our latest work:
  • Moving to nonlinear functions
Why bother with Linear Regression?

- Linear regression is important since many of recently developed machine learning techniques rely on this linearity property:
  - Support vector machines
  - Gaussian process regression
  - Radial basis functions
  - Kernel regression etc...

- Solutions to linear problems can be easily extended to nonlinear problems via locally weighted methods (e.g. Atkeson et al. 1997)
Start with the Simplest Scenario: Linear Regression

- Ordinary Least Squares (OLS) solution is:
  \[ b = \left( X^T X \right)^{-1} X^T y \]
  - Matrix inversion requires \( O(d^3) \) and \( X'X \) may be ill-conditioned!

- Our EM-based solution: Variational Bayesian Least Squares (NIPS 2005)
  - Handles high input dimensions
  - Detects features automatically
  - Is computationally efficient
Variational Bayesian Approximation

- We use a factorial variational approximation (Ghahramani & Beal 2001) to overcome intractability:

  - Factorize the posterior over hidden variables so that we can form a tractable lower bound on the marginal likelihood

\[
Q(\theta) = \prod_{i=1}^{p} Q_i(\theta_i) = Q(\alpha, b)Q(Z)
\]
Application: EMG Prediction from M1 Neuron Activity

- When reconstructing EMG traces:
  - Drastically improves computation time over traditional methods* (8 hrs vs. several weeks)
  - Achieves comparable performance to traditional methods*

*Traditional methods include LASSO regression, stepwise regression & a combinatorial-like model search using Least Mean Squares regression. Datasets from (Sergio & Kalaska 98; Kakei et al, 99)
Our Research

• Motivation:
  • Why create Bayesian black box systems?

• Past work:
  • The simplest scenario: linear regression
  • A more realistic scenario: noisy input data

• Our latest work:
  • Moving to nonlinear functions
Linear Regression with Noisy Input Data

- OLS gives biased estimates since it does not account for noise in input data.
Bayesian Linear Regression with Noisy Inputs

• Inspired by factor analysis (Ting et al, ICML 2006):

Again:
• Handles high input dimensions
• Detects features automatically
• Is computationally efficient

Moreover:
• Accurately identifies regression estimate in presence of noisy input data
Results: 10-70% Improvement for Strongly Noisy Data (SNR = 2)

Bayesian parameter estimation generalizes 10-70% better
Results: 7-50% Improvement for Less Noisy Data (SNR = 5)

Bayesian parameter estimation generalizes 7-50% better
Application: Parameter Identification in Robot Dynamics

Oculomotor Vision Head

- 10-20% better*
- 77 features

Anthropomorphic Arm

- 5-17% better*
- 110 features

*Better than ridge regression, LASSO regression, stepwise regression with additional project step to satisfy physical constraints of RBD parameters (Ting et al, RSS 2006)
Our Research

- **Motivation:**
  - Why create Bayesian black box systems?
- **Past work:**
  - The simplest scenario: linear regression
  - A more realistic scenario: noisy input data
- **Our latest work:**
  - *Moving to nonlinear functions*
Moving to Nonlinear Functions: Locally Weighted Regression (work in progress)

• Bayesian locally weighted regression:
  • Automatically determines size of neighborhood data contributing to a local model
Interested in a Faculty Position in Machine Learning?

- Contact Stefan Schaal:

  sschaal@usc.edu