Statistical Perspectives on Data Mining

- **Basics**
  - Inference = Estimation
  - Consistency (e.g. SLLN)
  - Uncertainty
  - Assumptions and Robustness
  - Model Averaging

- **Themes**
  - Classification
  - Clustering

Statistical Bottlenecks

- **Causal Inference: Most Common Sources of Error**
  - Absence of Causal Connection = Independence?
  - Latent Variables
  - Sample Selection Bias
  - Model Equivalence

- **Model Estimation versus Model Identification**

- **Statistical versus Practical Significance**

Statistical Viewpoints

- **Interpretability**
- **Characterize Uncertainty**
- **‘Strength Borrowing’**
- **Explicit Assumptions**
- **Regularization Principle**
  - Occam’s Razor

Statistical Methods

- Regression
- Polynomial Networks
- Discrimininant Analysis
- Projection Pursuit
- Decision Trees (CART)
- Radial Basis Functions
- Neural Networks

Statistical Methods (Contd.)

- Graphical Models (Bayesian networks)
- Non-parametric methods
- Statistical Learning Theory
- Principal Component Analysis
- Visualization: Sammon’s Mapping, Confusion Matrices