Lecture 29: 5/3/17

Recipe for Machine Learning

① Given data $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$

② a) Choose a decision function $h_{\vec{\theta}}(\vec{x}) = \dots$
parameterized by $\vec{\theta}$

b) Choose objective function $J_D(\vec{\theta}) = \dots$
relies on the data.

③ Learn by choosing parameters that optimize the objective $J(\theta)$

$$\hat{\vec{\theta}} = \arg\min_{\vec{\theta}} J_D(\vec{\theta})$$

④ Predict on new test example $\vec{x}_{new}$ using h

$$\hat{y} = h_\theta(\vec{x}_{new})$$

Decision Functions

- Linear Reg: $h_\theta(\vec{x}) = \vec{\theta}^T \vec{x}$ where $x_0 = 1$
- Disc. Models: $h_\theta(\vec{x}) = \arg\max_y p(y \mid \vec{x})$
- Gen. Model: $h_\theta(\vec{x}) = \arg\max_y p(\vec{x}, y)$

GDA
GMM

- Log. Reg: $p(y \mid \vec{x}) = \cancel{1/}[1 + \exp(-\theta^T x)]$
- Naive Bayes: $p(\vec{x}, y) = p(y) \prod_{m=1}^{M} p(x_m \mid y)$
- Perceptron: $h_\theta(\vec{x}) = \text{sign}(\theta^T x)$
- SVM: " "
- Kernel Perceptron: $h_\alpha(\vec{\theta}) = \dots$
- HMM: $h_\theta(\vec{x}) = \arg\max_{\vec{y}} p(\vec{x}, \vec{y})$
- NN: $p(y \mid \vec{x}) = \sigma(U^T \sigma(W^{(2)} \sigma(W^{(1)} \vec{x} + b^{(1)}) + b^{(2)}))$
- Matrix Factorization: $\hat{r}_{ij} = \vec{u}_i^T v_j$

Objective Functions

MLE: $J(\theta) = -\sum_{i=1}^{N} \log p(\vec{x}^{(i)}, y^{(i)})$

MCLE: $J(\theta) = -\sum_{i=1}^{N} \log p(y^{(i)} | \vec{x}^{(i)})$

MAP (joint): $J(\theta) = -\sum_{i=1}^{N} \log p(\vec{x}^{(i)}, y^{(i)}) - \log p(\theta)$

MAP (cond.): $J(\theta) = -\sum_{i=1}^{N} \log p(y^{(i)} | \vec{x}^{(i)}) - \log p(\theta)$

L2 Reg. : $J'(\theta) = J(\theta) + \lambda \|\theta\|_2^2$

L1 Reg : $+ \lambda \|\theta\|_1$

Marginal Likelihood: $J(\theta) = -\sum_{i=1}^{N} \log \sum_{z} p(\vec{x}, z)$

MF : $J(U, V)$

HMM :

SVM : (constraints)

Optimization Method

- Gradient Descent

while not conv.
$\vec{\theta} \leftarrow \vec{\theta} - \gamma \nabla J(\theta)$

- SGD

when $J(\theta) = \sum_{i=1}^{N} J_i(\theta)$

while not conv.
$i \sim \text{Uniform}(1, \ldots, N)$
$\vec{\theta} \leftarrow \vec{\theta} - \gamma \nabla J_i(\theta)$

- Closed Form
  1. Compute partial derivs.
  2. Set to zero and solve.
- Quadratic Prog.
- Coordinate Descent
- Block Coordinate Descent.