**【CS】Machine Learning Problem Set**

**11 April 2018**

Undergraduate level theoretical machine learning problems.

**1. Bias-Variance Decomposition and Cross Validation**

1. 将期望泛化误差分解为bias，variance和noise
2. k-fold CV中k取值越大，bias，variance如何变化？LOO相对于k-fold有什么优劣？

**2. Linear Regression**

1. 一元线性回归模型f(x)=wTx+bf(x)=wTx+b，通过最小化均方误差推导参数的闭式解
2. 多元线性回归模型f(X)=WTXf(X)=WTX，通过最小化均方误差推导参数的闭式解
3. 证明对线性回归模型做牛顿法的更新可以得到其最优解
4. 线性回归的概率模型y|x,w∼N(wx,σ2)y|x,w∼N(wx,σ2)，通过最大化似然函数求闭式解

**3. Logistic Regression**

1. 假设lnp(y=1|x)p(y=0|x)=wTx+bln⁡p(y=1|x)p(y=0|x)=wTx+b，写出在一个数据集上的log-likelihood function
2. 通过最大化likelihood function来推导其梯度下降算法的公式

**4. Decision Tree**

1. 给定一个数据集，基于信息增益找到对于数据集进行划分的特征（计算题）
2. 推导gradient boosting tree每个节点上的增益计算公式和权重更新公式

**5. Neural Network**

推导BP算法公式

**6. SVM**

1. 给出线性可分SVM原问题，写出其对偶性形式；样本点是支持向量的条件
2. 给出线性不可分（soft-margin）SVM原问题，写出其对偶形式；样本点在margin上/内/外的条件
3. 简要概述一种有效解决SVM的算法（SMO）
4. 写出核化SVM以及核化线性回归的形式

**7. MAP and ML**

1. 给定一个样本序列，用MAP和ML求相应的参数
2. 如果使用full Bayesian方法应该怎么求？Full Bayesian方法有什么缺点？

**8. Bayesian Network**

1. 给定一个贝叶斯网络，给定某个随机变量的观测值，求另一个随机变量的概率分布
2. 给定一个贝叶斯网络，当某个随机变量给定的时候求另外两个随机变量是否独立（可以公式写出说明，也可以使用moralization方法）
3. 给定一个贝叶斯网络，给定证据变量和待查询变量，用Gibbs sampling求待查询变量的概率分布；根据Markov Blanket求解待改变变量的概率分布

**9. EM Algorithm**

1. 通过对J(c,μ)=∑Ni=1||x(i)−μc(i)||2J(c,μ)=∑i=1N||x(i)−μc(i)||2做coordinate descent，推导k-mean参数更新公式
2. 推导Gaussian mixture model的EM算法迭代公式；GMM与k-means有什么关系？
3. 推导一般问题的EM算法迭代公式
4. 证明EM算法的收敛性

**10. Ensemble**

1. 证明相互独立的N个分类器进行majority voting，错误率随N指数下降（Hoeffding不等式）
2. 【AdaBoost】简述什么是additive model；通过最小化指数损失函数L=Ex∼D[e−yf(x)]L=Ex∼D[e−yf(x)]推导AdaBoost的样本权重更新公式
3. 证明bagging能够减小variance；简述随机森林模型

**11. Clustering**

层次聚类（hierarchical clustering）的几种距离度量有什么优劣（min，max，mean）？层次聚类结果的表示方法（dendrogram）；相比于k-mean的优劣。

**12. kNN**

假设对于任意x和小正数e，在x附近e距离范围内都能找到一个训练样本，证明kNN的泛化错误率不超过贝叶斯最优分类器错误率的两倍

**13. PCA**

1. 说明PCA特征值分解的方法等价于找到某个向低维空间的投影，使得各个样本的重构误差最小
2. 说明PCA特征值分解的方法等价于找到某个向低维空间的投影，使得各个样本投影的方差最大

**14. PAC and VC Dimension**

1. 证明样本数足够大的时候，任意一个假设的在样本上的经验误差与泛化误差大概率接近；
2. 证明样本数足够大的时候，通过最小化经验误差得到假设的泛化误差与假设空间中最小泛化误差大概率接近；
3. 求假设的VC维
   * h(x)=I(a<x)h(x)=I(a<x)
   * h(x)=I(a<x<b)h(x)=I(a<x<b)
   * h(x)=I(asinx>0)h(x)=I(asin⁡x>0)
   * h(x)=I(sin(x+a)>0)h(x)=I(sin⁡(x+a)>0)

**15. Probabilistic Graphical Model**

1. 写出隐马可夫模型的联合概率分布，并使用该公式计算某个序列的概率
2. 给定一个马可夫条件场，以及相应的势函数，计算联合概率和条件概率

**16. Searching**

1. BFS、DFS、A\*算法complete和optimal的条件，它们的时间和空间复杂度
2. 证明满足h(n)≥c(n,n′)+h(n′)h(n)≥c(n,n′)+h(n′)的A\*算法能找到最优解
3. 简述模拟退火和遗传算法

**17. Guassian Discriminate Analysis**

yx∥y=0x∥y=1∼Bernoulli(Φ)=p(y)=Φy(1−Φ)1−y∼N(μ\_0,Σ)=1(2π)n/2∥Σ∥1/2exp(−12(x−μ\_0)TΣ−1(x−μ\_0))∼N(μ\_1,Σ)y∼Bernoulli(Φ)=p(y)=Φy(1−Φ)1−yx‖y=0∼N(μ\_0,Σ)=1(2π)n/2‖Σ‖1/2exp⁡(−12(x−μ\_0)TΣ−1(x−μ\_0))x‖y=1∼N(μ\_1,Σ)

通过最大化以上模型的似然函数求其参数估计

**18. Naive Bayes**

1. 对于垃圾邮件判别任务写出如何利用朴素贝叶斯模型来做判断
2. 通过最大化似然函数来推导参数估计公式

**19. Online Learning**

证明perceptron algorithm所犯错误次数有上界