机器学习第三讲习题以及课外阅读

习题 （任选两小题作答）

问题1. 逻辑回归交叉熵损失函数

吴恩达视频讲义中定义了每个样本点(x,y)的概率估计损失函数为:

$$cost(h_\theta(x),y) = \begin{cases} -\log(h_\theta(x)), & y=1; \\ -\log(1-h_\theta(x)), & y=0. \end{cases}$$

试证明:

$cost(h_\theta(x),y) = -[y\cdot\log(h_\theta(x)) + (1-y)\cdot\log(1-h_\theta(x))]$,

右端表达式称为逻辑回归的交叉熵损失函数。

问题2. 逻辑回归极大似然解释

假设$P(Y=1|x;\theta) = h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$,

试用极大似然估计法推导吴恩达视频讲义中的损失函数(不含正则项)。

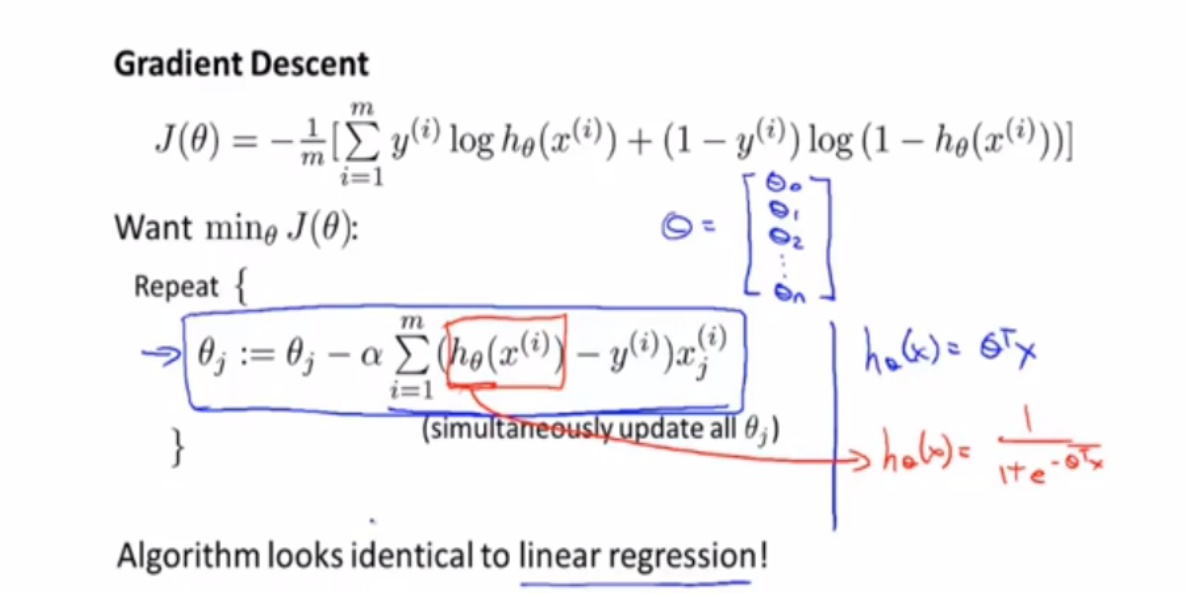
问题3. 判别模型与生成模型

请查阅相关资料，简单描述判别模型与生成模型，并指出逻辑回归和朴素贝叶斯分类分别属于哪一类模型。

问题4. 试用softmax函数将二分类推广为多分类并尽可能描述细节(损失函数、梯度等)。

课外阅读

吴恩达视频讲义中出现了如下梯度公式:



2.1 逻辑回归损失函数偏导数推导.

$J(\theta) = \frac{-1}{m}[\sum_{i=1}^m y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$

求偏导数 $\frac{\partial}{\partial\theta_j}J(\theta)$, 以及梯度 $\nabla J(\theta)$.

解答:

Let $x^{(i)} = (x_0^{(i)}=1, x_1^{(i)}, x_2^{(i)}, ..., x_k^{(i)}, ..., x_n^{(i)})^T$, $\theta = (\theta_0,\theta_1,\theta_2,...,\theta_k,...,\theta_n)^T$ and $h_\theta(x^{(i)}) = \frac{1}{1+e^{-\theta^T x^{(i)}}}$. Then we have

$$h_\theta(x^{(i)}) = \frac{1}{1+e^{-\sum_{j=0}^{n}\theta_j \cdot x_j^{(i)}}}.$$

Defining $z = -\sum_{j=0}^{n}\theta_j \cdot x_j^{(i)}$, we have $h_\theta(x^{(i)}) = \frac{1}{1+e^z}$.

Thus, by chain rule about taking partial derivatives, we get

$\frac{\partial}{\partial\theta_j}J(\theta) = \frac{\partial}{\partial\theta_j}\frac{-1}{m}[\sum_{i=1}^{m} y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$
$= \frac{\partial}{\partial\theta_j}\frac{-1}{m}[\sum_{i=1}^{m} y^{(i)}\log\frac{1}{1+e^z} + (1-y^{(i)})\log(1-\frac{1}{1+e^z})]$
$= \frac{-1}{m}[\sum_{i=1}^{m}(y^{(i)}\frac{d}{dz}\{\log\frac{1}{1+e^z}\}\cdot\frac{\partial z}{\partial\theta_j} + (1-y^{(i)})\frac{d}{dz}\{\log\frac{1}{1+e^{-z}}\}\cdot\frac{\partial z}{\partial\theta_j})]$
$= \frac{-1}{m}[\sum_{i=1}^{m}(y^{(i)}\frac{d}{dz}\{-\log(1+e^z)\} + (1-y^{(i)})\frac{d}{dz}\{-\log(1+e^{-z})\})\cdot\frac{\partial z}{\partial\theta_j}]$
$= \frac{-1}{m}[\sum_{i=1}^{m}(y^{(i)}\cdot\frac{-e^z}{1+e^z} + (1-y^{(i)})\cdot\frac{e^{-z}}{1+e^{-z}})\cdot\frac{\partial z}{\partial\theta_j}]$
$= \frac{-1}{m}[\sum_{i=1}^{m}(y^{(i)}\cdot(h_\theta(x^{(i)})-1) + (1-y^{(i)})\cdot h_\theta(x^{(i)}))\cdot(-x_j^{(i)}))]$
$= \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)})-y^{(i)})\cdot x_j^{(i)}$

进一步的, $\nabla J(\theta) = (\frac{\partial}{\partial\theta_0}J(\theta), \frac{\partial}{\partial\theta_1}J(\theta), ..., \frac{\partial}{\partial\theta_n}J(\theta))^T$
$= \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)})-y^{(i)})x^{(i)}$
= $E[(h_\theta(X)-Y)X]$, 需要注意的是这里 X 是空间 R^{n+1} 里的向量。

沿着$\nabla J(\theta)$方向，损失函数增加的最快; 沿着$-\nabla J(\theta)$方向，损失函数减少的最快。

2.2 逻辑回归函数的正则化.

考虑如下正则化后的损失函数:
$J(\theta) = \frac{-1}{m}[\sum_{i=1}^{m} y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$.
则偏导数 $\frac{\partial}{\partial\theta_j}J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)})-y^{(i)})\cdot x_j^{(i)} + I\{j\geq 1\}\cdot\frac{\lambda}{m}\theta_j$,
梯度$\nabla J(\theta) = E[(h_\theta(X)-Y)X] + \tilde{\theta}$, 其中$\tilde{\theta} = (0,\theta_1,\theta_2,...,\theta_n)^T$.

2.3 分类问题算法代码分享(待更新).