MLE & MAP

1. 最大似然估计（MLE）：
   定义: 最大似然估计是一种参数估计方法，它选择能够使给定观测数据出现概率（似然函数）最大的参数值。
   公式:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta; X)$$

   其中，$L(\theta; X)$ 是似然函数，θ是模型参数，X 是观测数据。
2. 最大后验概率估计（MAP）：
   定义: 最大后验概率估计结合了先验概率和似然函数，选择能够使参数的后验概率最大化的参数值。
   公式:

$$\hat{\theta}_{MAP} = \arg\max_{\theta} (P(\theta|X)) = \arg\max_{\theta} (P(X|\theta)P(\theta))$$

   其中，$P(\theta|X)$ 是后验概率， $P(X|\theta)$是似然函数，$P(\theta)$是关于参数的先验概率分布。

实例

IV. [8 points] Maximum Likelihood Estimate (MLE)

You are given a dataset with N records in which the nth record has an input attribute x[n], which is generated from a Gaussian distribution as following:

$$p(\mathbf{x}|\,A) = \frac{1}{(2\pi A)^{N/2}} \exp\left(-\frac{1}{2A}\sum_{n=0}^{N-1}(x[n]-A)^2\right).$$

We have one unknown parameter A in the above model and we want to learn the MLE of A from the data.

注意这里给定的条件就已经是最大似然函数了，不需要再拆开额外算一遍，因为x是一个数据集，也是一个数组，那么$p(x|A) = \prod_{n=0}^{N-1} p(x[n]|A)$

直接对上面的式子取对数后求导即可，算出来的就是使得A最大的各参数函数。

MAP

注意MLE算的是p(x|A)，而MAP算的是后验概率p(A|x)，后者算的是后验概率，由贝叶斯公式得

$p(A|x) = \frac{p(x|A) \cdot p(A)}{p(x)}$

因为p(x) = 1，则这里可以忽略

$p(A|x) = p(x|A) \cdot p(A)$

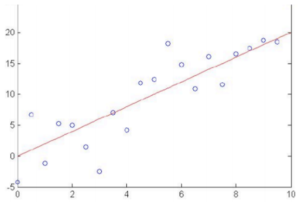
$p(A)$称为先验概率，$p(A|x)$称为后验概率

总的来说，p(A)在进行实验之前，就可以根据经验或者分析来获得，是属于先验概率，而$p(A|x)$是在实验发生后再观测讨论A发生与x之间的概率联系。

考试的话如果出了就记住 后验概率 = 极大似然函数 × 先验概率

MCLE

Training Linear Regression



$$p(y|x;W) = N(w_0 + w_1 x, \sigma)$$

How can we learn W from the training data?

Learn Maximum Conditional Likelihood Estimate!

$$W_{MCLE} = \arg\max_W \prod_l p(y^l|x^l, W)$$

$$W_{MCLE} = \arg\max_W \sum_l \ln p(y^l|x^l, W)$$

where

$$p(y|x;W) = \frac{1}{\sqrt{2\pi\sigma^2}} \quad e^{-\frac{1}{2}(\frac{y-f(x;W)}{\sigma})^2}$$

条件最大似然估计用于评估给定线性回归拟合函数中的回归直线估计，包括斜率和截距的估计，注意这里是对W求导

和之前的极大似然估计其实关系并不大

在上图的上下文中，W 指的是线性回归模型的参数，它通常包括权重和截距。在数学表达式 $p(y|x;W) = N(w_0 + w_1 x, \sigma)$ 中，W 表示模型的参数集合，这里可能是指 w_0（截距项）和 w_1（斜率或权重）。N 表示正态分布，σ 是标准差，表示预测中的不确定性或噪声。

模型试图通过调整参数 W 来最好地拟合训练数据，方法是最大化条件似然函数，这在数学上表达为 $W_{\text{MCLE}} = \arg\max_W \prod_l p(y^l|x^l, W)$，其中 l 表示不同的数据点。

在实践中，为了方便计算，通常会最大化似然函数的对数形式，因为乘积的对数可以转换为和的形式，即 $W_{\text{MCLE}} = \arg\max_W \sum_l \ln p(y^l|x^l, W)$。

所以在这个上下文中，W 是我们试图通过最大化条件似然估计来学习的线性回归模型的参数。