初识机器学习

韩雅妮

XiDian University

目录

1. 机器学习的相关介绍
2. 单变量线性回归
3. 梯度下降算法
4. 多变量线性回归
5. 正规方程法

机器学习

实现人工智能的一个很好的方法就是采用学习算法模拟人类大脑的学习方式。机器学习实际上是从人工智能发展出来的一个领域，让机器具有学习能力的研究领域。

机器学习

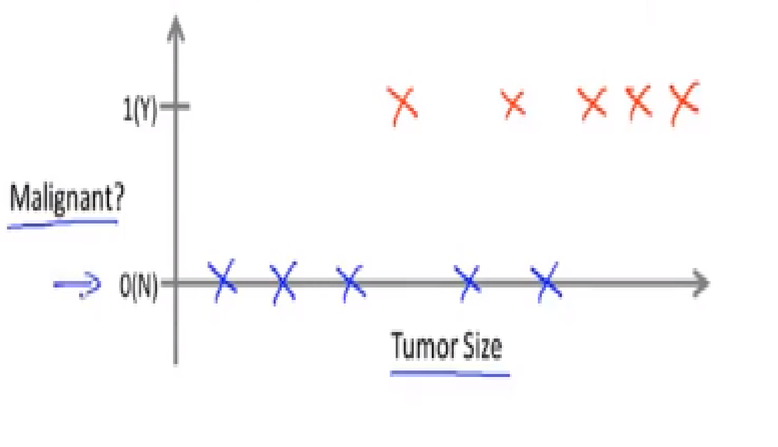
如果一个程序可以在任务T上，随着经验E的增加，效果P也会增加，则这个程序可以在经验中学习。

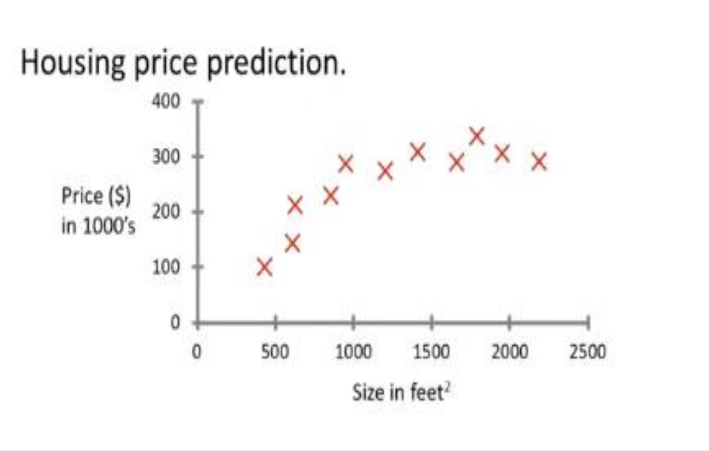
机器学习已经应用到了很多领域中，比如使用机器学习算法来挖掘数据；通过医疗记录来了解病情；计算生物学等。

监督学习

监督学习：训练样本中包含有特征和标签信息。

典型的问题就是分类问题(Classfication)和回归问题(Regression)

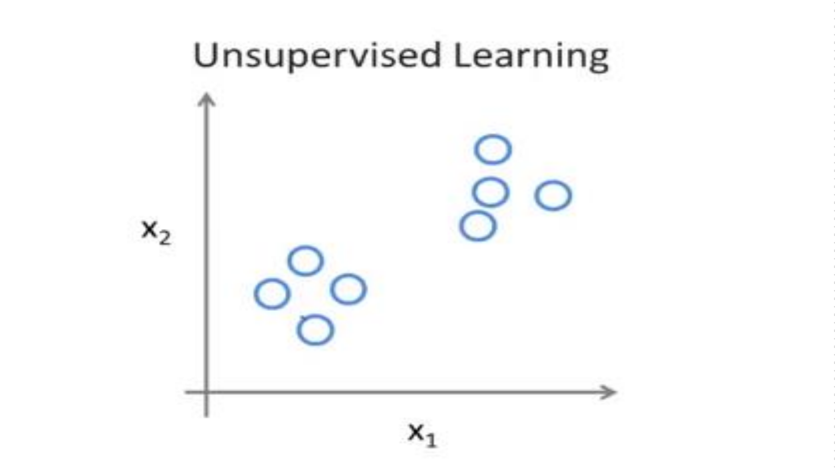




无监督学习

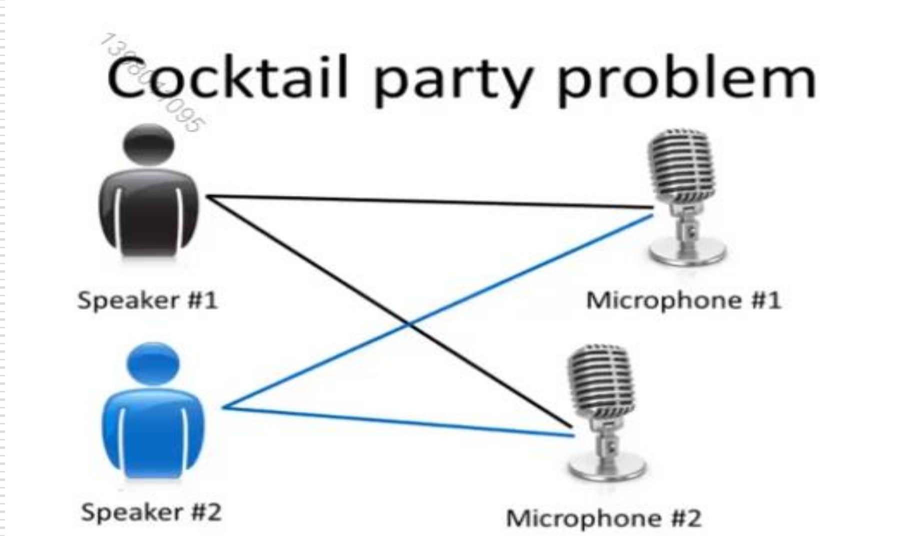
无监督学习：没有标签，给定数据集。

如聚类算法。可以通过该算法对顾客进行划分，将他们划分成不同的类，从而为同一类中的顾客进行产品推荐。



鸡尾酒会问题：

假设酒会有两个人，每个人与麦克风的距离不同。

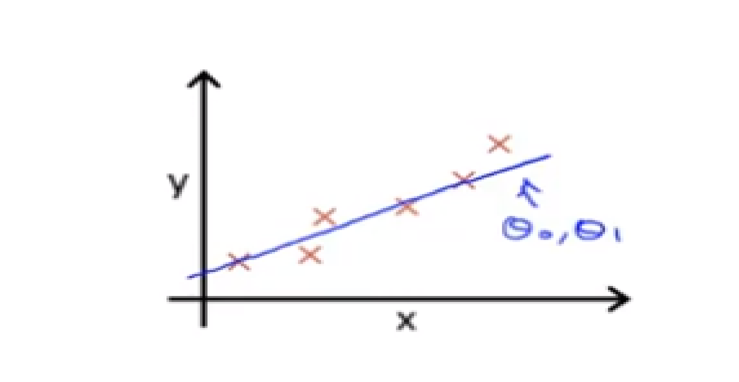


研究人员记录麦克风中的声音，利用无监督算法可以找出数据的结构，分离出两个被叠加到一起的混合音频源。

单变量线性回归

假设函数 $h_{\boldsymbol{\theta}}(x) = \boldsymbol{\theta_0} + \boldsymbol{\theta_1} x$

$\boldsymbol{\theta_0}$和$\boldsymbol{\theta_1}$是模型参数，希望选择合适的参数，能够通过x预测出y。



线性回归的目标函数：$\min\limits_{\boldsymbol{\theta_0}\boldsymbol{\theta_1}} \frac{1}{2m} \sum_{i=1}^{m}(h_{\boldsymbol{\theta}}(x) - \boldsymbol{y})^2$

$$J(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \frac{1}{2m} \sum_{i=1}^{m}(h_{\boldsymbol{\theta}}(x) - \boldsymbol{y})^2$$

代价函数(平方误差函数)：$\min\limits_{\boldsymbol{\theta_0}\boldsymbol{\theta_1}} J(\boldsymbol{\theta_0}, \boldsymbol{\theta_1})$

梯度下降算法：

$$\theta_{\mathrm{j}} := \theta_{\mathrm{j}} - \alpha\frac{\partial}{\partial\theta_{\mathrm{j}}} \mathrm{J}(\theta_0, \theta_1)$$

其中 α：表示梯度下降的幅度，学习率；

θ_0，θ_1要同步更新。

使用梯度下降算法最小化平方差代价函数

$$\frac{\partial}{\partial\theta_{\mathrm{j}}} \mathrm{J}(\theta_0, \theta_1) = \frac{\partial}{\partial\theta_{\mathrm{j}}} \frac{1}{2m} \sum_{i=1}^{m}(\mathrm{h}_\theta \mathrm{x}^{(i)} - y^{(i)})^2$$

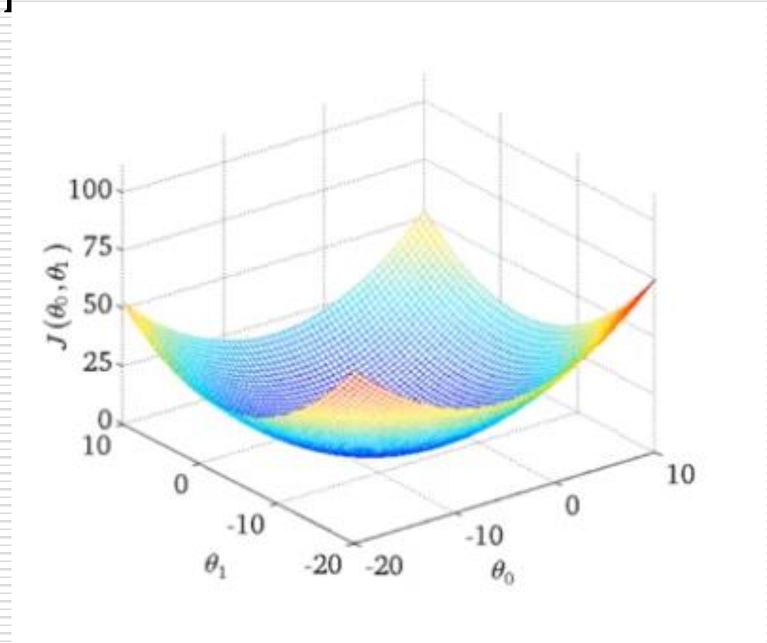
$$= \frac{\partial}{\partial\theta_{\mathrm{j}}} \frac{1}{2m} \sum_{i=1}^{m}(\theta_0+\theta_1\mathrm{x}^{(i)} - y^{(i)})^2$$

当$j=0$时： $\frac{\partial}{\partial \boldsymbol{\theta}_0} \mathrm{J}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \frac{1}{\mathrm{m}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} (\mathrm{h}_{\boldsymbol{\theta}} \mathrm{x}^{(\boldsymbol{i})} - \boldsymbol{y}^{(\boldsymbol{i})})$

当$j=1$时： $\frac{\partial}{\partial \boldsymbol{\theta_1}} \mathrm{J}(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}) = \frac{1}{\mathrm{m}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} (\mathrm{h}_{\boldsymbol{\theta}} \mathrm{x}^{(\boldsymbol{i})} - \boldsymbol{y}^{(\boldsymbol{i})}) \mathrm{x}^{(\boldsymbol{i})}$

$$\boldsymbol{\theta_0} := \boldsymbol{\theta_0} - \boldsymbol{\alpha} \frac{1}{\mathrm{m}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} (\mathrm{h}_{\boldsymbol{\theta}} \mathrm{x}^{(\boldsymbol{i})} - \boldsymbol{y}^{(\boldsymbol{i})})$$

$$\boldsymbol{\theta_1} := \boldsymbol{\theta_1} - \boldsymbol{\alpha} \frac{1}{\mathrm{m}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} (\mathrm{h}_{\boldsymbol{\theta}} \mathrm{x}^{(\boldsymbol{i})} - \boldsymbol{y}^{(\boldsymbol{i})}) \mathrm{x}^{(\boldsymbol{i})}$$



多变量线性回归

多个特征值 $h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$

代价函数： $J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_{\theta}x^{(i)} - y^{(i)})^2$

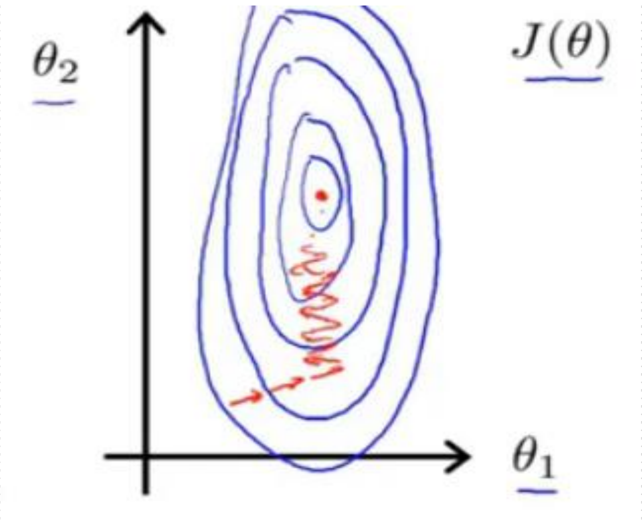
其中θ为n+1维向量

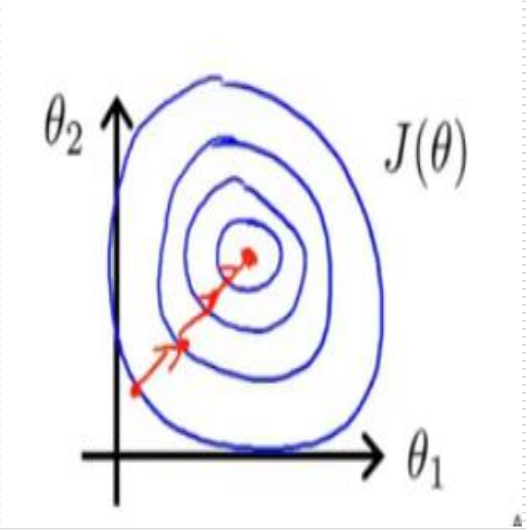
梯度下降： $\theta_j := \theta_j - \alpha\frac{\partial}{\partial\theta_j} J(\theta)$

$$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_{\theta}x^{(i)} - y^{(i)})\, x_j^{(i)}$$

$\boldsymbol{\alpha}$的选择：不断的尝试，绘制出J($\boldsymbol{\theta}$)的图像，选择使J($\boldsymbol{\theta}$)快速下降的$\boldsymbol{\theta}$，确保梯度下降算法正常进行。

特征缩放：将特征值缩放到相似范围之内





正规方程法

提供了一种求解θ的解析解法，可一次性求解θ的最优值。

如何最小化代价函数：

$$\frac{\partial}{\partial \theta_j} \mathrm{J}(\boldsymbol{\theta}) = \cdots = 0 \ ,$$

从而求出$\boldsymbol{\theta_0}$，$\boldsymbol{\theta_1}$，$\cdots$ ，$\boldsymbol{\theta_n}$

正规方程法

假设有m个样本：$(x^{(1)}, y^{(1)}),\ldots, (x^{(m)}, y^{(m)})$

n个特征向量： $x^{(i)} = \begin{bmatrix} x_1^{(i)} \\ \ldots \\ x_n^{(i)} \end{bmatrix}$

构建 $x = \begin{bmatrix} (x^{(1)})^T \\ \ldots \\ (x^{(m)})^T \end{bmatrix}$, $y = \begin{bmatrix} y^{(1)} \\ \ldots \\ y^{(m)} \end{bmatrix}$

则：$J(\boldsymbol{\theta}) = \frac{\mathbf{1}}{\mathbf{2m}}(\boldsymbol{x}\theta - \boldsymbol{y})^{\mathbf{2}}$

$$= \frac{\mathbf{1}}{\mathbf{2m}}(\boldsymbol{x}\theta - \boldsymbol{y})^{\boldsymbol{T}}\ (\boldsymbol{x}\theta - \boldsymbol{y})$$

$$= \frac{\mathbf{1}}{\mathbf{2m}}(y^{T}\mathrm{y} - y^{T}\boldsymbol{x}\theta - \theta^{T}x^{T}\mathrm{y} + \theta^{T}x^{T}\boldsymbol{x}\theta)$$

$$2x^{T}\mathrm{y} = 2x^{T}\boldsymbol{x}\theta$$

所以： $\theta = (x^{T}x)^{-1}x^{T}y$

注意：如果遇到不可逆方程时，检查是否有无用特征，是否包含太多特征。

梯度下降法：

缺点：需要多次尝试，选择α；

多次迭代计算参数，速度慢。

优点：当特征向量非常大时，仍然能正常工作。

正规方程法：

优点：不需选择α，不用迭代，一次完成计算。

缺点：当特征向量的维数n比较大时，$(x^T x)^{-1}$的计算会变慢。

Thanks