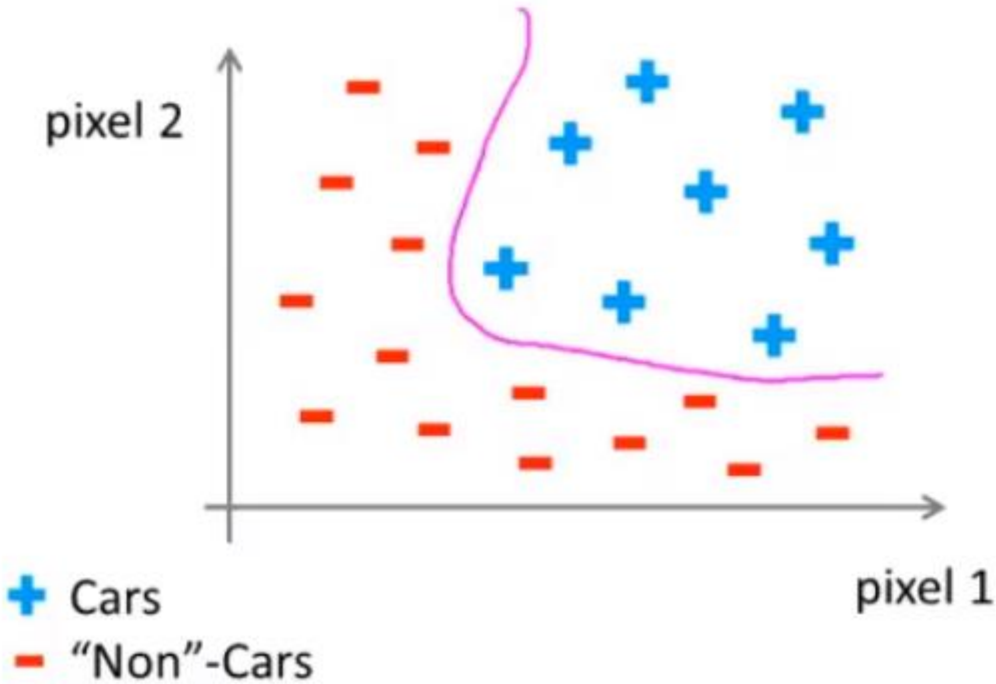
1、非线性假设

之前所学的房价预测模型，通常只有少量几个特征值，根据特征值，使用 logistic 回归模型得到了较好的拟合决策边界。但是，当特征值数量变多后，且分类图像中，并无明显区分边界时，此时我们考虑使用更多特征值。且增加次数的办法，以此得到较好的拟合边界。譬如房价预测中，我们含有 100 个特征值，如果我们要包含所有高次项，显然会使得特征向量急剧膨胀。因此，当特征数量很多时，一昧地增加次数项不是一个明智的做法。

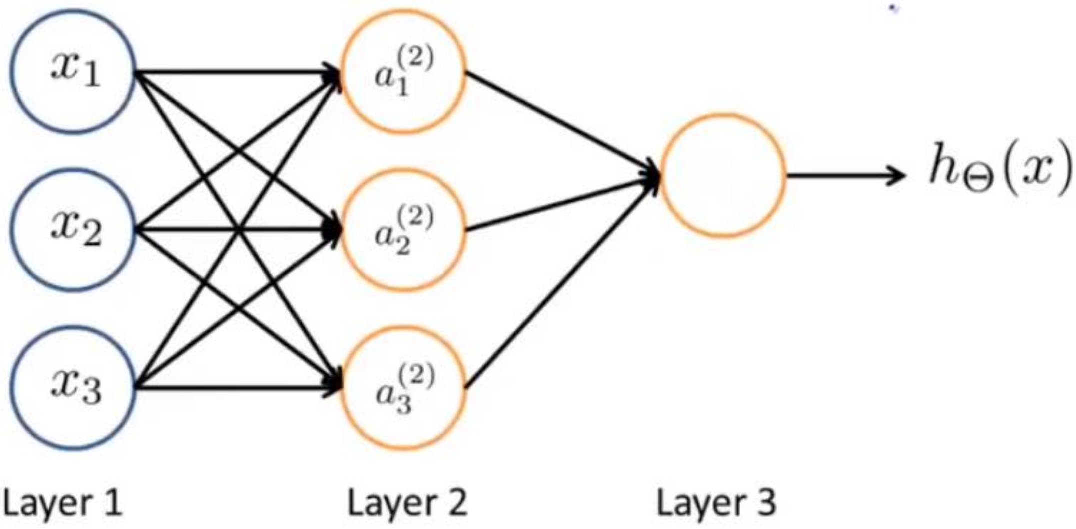
$$
\begin{aligned}
x_1 &= \text{size} \\
x_2 &= \text{\# bedrooms} \\
x_3 &= \text{\# floors} \\
x_4 &= \text{age} \\
&\cdots \\
x_{100}
\end{aligned}
$$

再来看另一个关于图像识别的例子，加入你有很多张包含汽车的图片，每张图片像素为 50*50，先将数据可视化出来，这里横、纵坐标轴只是将特征简化了，实际上每



个样本有 2500 个特征值，根本无法用图形表示出来，如果是 RGB 图像，那会有 7500 个特征值，如果还要通过包含 2 次项来拟合出非线性假设，那么总的特征值将超过 300 万个！根本无法再用于 logistic 回归假设了。

2、神经网络初窥



在这样一幅图中，Layer1 称为输入层，Layer2 称为隐藏层，Layer3 称为输出层。用$a_i^{(j)}$表示第 j 层第 i 个处理单元（输入，隐藏），Θ^j表示从 j 层到 j+1 层的权重矩阵（参数矩阵）。如下就是每个层的输入与输出对应关系。

$$
\begin{aligned}
a_1^{(2)} &= g(\Theta_{10}^{(1)}x_0 + \Theta_{11}^{(1)}x_1 + \Theta_{12}^{(1)}x_2 + \Theta_{13}^{(1)}x_3) \\
a_2^{(2)} &= g(\Theta_{20}^{(1)}x_0 + \Theta_{21}^{(1)}x_1 + \Theta_{22}^{(1)}x_2 + \Theta_{23}^{(1)}x_3) \\
a_3^{(2)} &= g(\Theta_{30}^{(1)}x_0 + \Theta_{31}^{(1)}x_1 + \Theta_{32}^{(1)}x_2 + \Theta_{33}^{(1)}x_3)
\end{aligned}
$$

$$
h_\Theta(x) = a_1^{(3)} = g(\Theta_{10}^{(2)}a_0^{(2)} + \Theta_{11}^{(2)}a_1^{(2)} + \Theta_{12}^{(2)}a_2^{(2)} + \Theta_{13}^{(2)}a_3^{(2)})
$$

Θ^1 表示从第一层到第二层的权重矩阵，由于第二层有 3 个单元，第一层有 3 个单元，因此Θ^1是一个 3*（3+1）的矩阵。

由输入层到隐藏层再到输出层逐步计算激活函数的过程我们称之为前向传播。

神经网络同样可以用来处理多分类问题，这里依然采用逻辑分类的思想。首先比较代价函数的不同

一对二分类中代价函数：

$$
J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} log\left(h_\theta\left(x^{(i)}\right)\right) + (1 - y^{(i)}) log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2
$$

一对多分类中代价函数：

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{k} {y_k}^{(i)} log\left(h_\theta\left(x^{(i)}\right)\right)_k + (1-{y_k}^{(i)}) log\left(1-h_\theta\left(x^{(i)}\right)_k\right)\right]$$

$$+\frac{\lambda}{2m}\sum_{L=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(\Theta_{ji}^{(l)}\right)^2$$