

监督学习： (1 or 0)

1.回归 (regression)：指我们的目标是预测一个连续的输出值。目标是预测离散值输出。

2.分类 (classification)：

无监督学习： (归类)

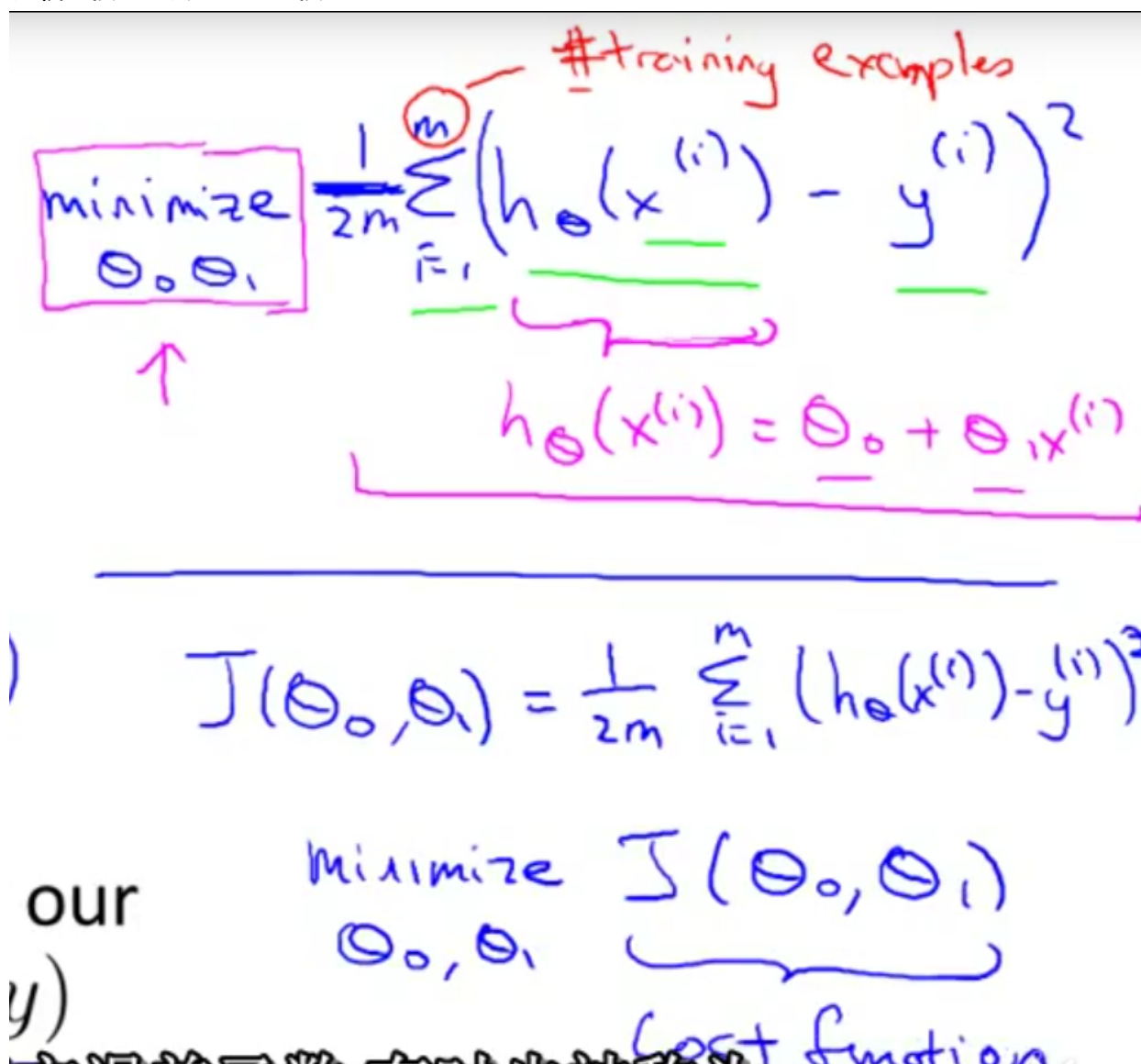
1.聚类算法：市场分割分类客户 社交群体分类组员

2.鸡尾酒算法：

Octave软件

线性回归 (linear regression) 模型：

代价函数：平方误差函数



The image shows handwritten notes on a piece of paper. At the top, it says "minimize θ_0, θ_1 " in a pink box. An arrow points from this box to the cost function formula below. The formula is
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$
 with several annotations: a red circle around the 'm' in the denominator, a red line pointing to it with the text "#training examples", green underlines under $h_{\theta}(x^{(i)})$ and $y^{(i)}$, and a pink bracket under the entire sum term. Below the formula, the hypothesis is written as $h_{\theta}(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$ in pink. A horizontal blue line separates this from the next part. Below the line, the cost function is written again as $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$. At the bottom, it says "Minimize $J(\theta_0, \theta_1)$ " in blue, with a blue bracket under $J(\theta_0, \theta_1)$ and the words "Cost function" written in blue below it. On the left side, the words "our" and "y)" are partially visible.

minimize θ_0, θ_1

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

$h_{\theta}(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$

Minimize $J(\theta_0, \theta_1)$

Cost function

Hypothesis:

$$\underline{h_{\theta}(x) = \theta_0 + \theta_1 x}$$

Parameters:

$$\underline{\theta_0, \theta_1}$$



Cost Function:

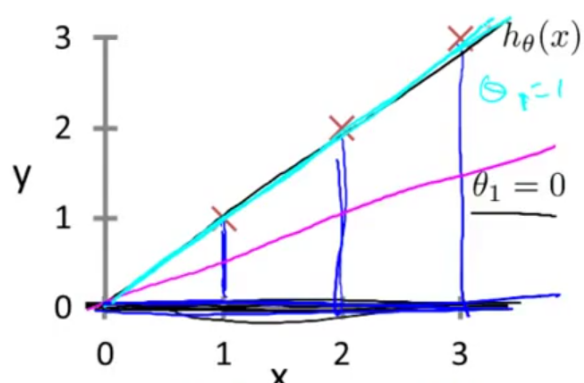
$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Goal: minimize $J(\theta_0, \theta_1)$
 θ_0, θ_1

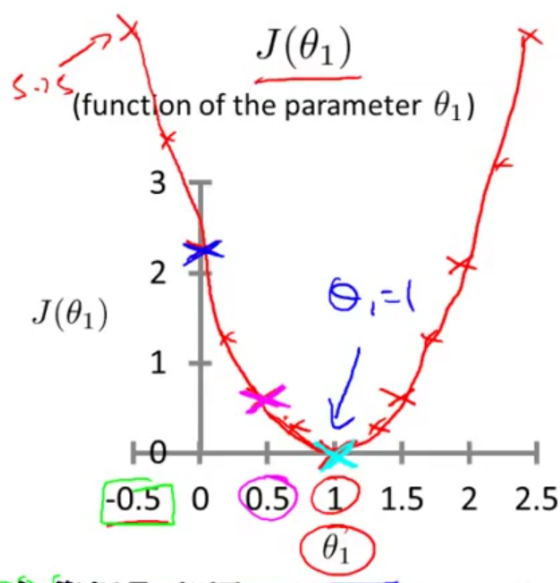
而且通过选择不同的参数 我们会得到不同的直线拟合

AcceptedDoge Bilibili $h_{\theta}(x)$

(for fixed θ_1 , this is a function of x)



$$J(\theta) = \frac{1}{2m} (1^2 + 2^2 + 3^2)$$



梯度下降:

Gradient descent algorithm

repeat until convergence {
→ $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ (for $j = 0$ and $j = 1$)
}

Correct: Simultaneous update

$\text{temp0} := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$

$\text{temp1} := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$

$\theta_0 := \text{temp0}$

$\theta_1 := \text{temp1}$

让我们来看看这个公式有很多细节问题

α -learning rate: 以多大的幅度更新参数 θ_j , $j=0, 1$

导数项: 下降斜率

α 过小, 下降过慢, α 过大下降过快会错过最低点。