

Machine Learning Notes

beck

2021 年 1 月 7 日

Week1

1. Classification problem: discrete valued output. $(0, 1, 2 \dots)$
2. Regression problem: to predict a continuous valued output.
3. Supervised learning: w/ correct answer.
4. Unsupervised learning: w/o correct answer, to find structure in data.
5. Linear regression: fit a straight line. (hypothesis function is linear, e.g. $H(x) = Ax + B$)
6. Hypothesis function: maps input x to output y .

$$H(x) = y$$

7. Cost function: choose model parameters θ_i .

$$J(\theta_0, \theta_1, \dots, \theta_i) \ (i = 0, 1, \dots)$$

8. Linear regression cost function: mean squared error(MSE).

$$J(\theta_i) = \frac{1}{m} \sum_{i=1}^m (h(x_i) - y_i)^2$$

9. Contour plots: 等高线图
10. Converge/Diverge: 收敛/发散
11. Gradient descent algorithm: minimize cost function.

$$\begin{aligned} & \text{repeat until convergence} \{ \\ & \theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1, \dots, \theta_j) \\ & \} \end{aligned}$$

12. Gradient descent simultaneous update:

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

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

...

$$tempj = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1, \dots, \theta_j)$$

$$\theta_0 = temp0$$

$$\theta_1 = temp1$$

...

$$\theta_j = tempj$$

13. No need to decrease α over time, because gradient descent will automatically take smaller steps. ($|\frac{\partial}{\partial \theta} J(\theta)|$ decreases as approaching local minimum, and finally turn 0)
14. Batch gradient descent: each step of gradient descent uses all the training examples.
15. For the specific choice of cost function $J(\theta)$ used in linear regression, there are no local optima (other than the global optimum) (optimum, plural noun: optima)