Machine Learning Formulas

MingShun Wu

October 6, 2016

Contents

1 线性回归(Linear Regression)

1.1 各向量形式

在机器学习中，各个变量在单独出现时均为列向量的形式，但若是以多个向量形成的矩阵形式出现时，均为行向量的形式。
以下将会写出各个变量单独出现的情况、各变量以矩阵的形式一起出现的情况。

1. $\vec{x}$

$$\vec{x} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}_{(n+1)*1} \tag{1}$$

2. $\vec{\theta}$

$$\vec{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}_{(n+1)*1} \tag{2}$$

3. $\vec{y}$

$$\vec{y} = \begin{pmatrix} y \end{pmatrix}_{1*1} \tag{3}$$

1.2 批梯度下降

1.2.1 各矩阵形式

下述式子中均为矩阵的形式

1. X

$$\begin{aligned} X &= \begin{pmatrix} \vec{x}^{(1)} \\ \vec{x}^{(2)} \\ \vec{x}^{(3)} \\ \vdots \\ \vec{x}^{(m)} \end{pmatrix} \\ &= \begin{pmatrix} x_0^{(1)} & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_n^{(1)} \\ x_0^{(2)} & x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_n^{(2)} \\ x_0^{(3)} & x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_0^{(m)} & x_1^{(m)} & x_2^{(m)} & x_3^{(m)} & \dots & x_n^{(m)} \end{pmatrix} \\ &= \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & x_3^{(m)} & \dots & x_n^{(m)} \end{pmatrix}_{m*(n+1)} \end{aligned} \tag{4}$$

2. $\vec{\theta}$

$$\vec{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_n \end{pmatrix}_{(n+1)*1} \tag{5}$$

3. $h_{\vec{\theta}}(\vec{x})$

$$\begin{aligned} h_{\vec{\theta}}(\vec{x}) &= X * \vec{\theta} \\ &= \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_n^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_n^{(2)} \\ 1 & x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & \dots & x_n^{(3)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} & x_3^{(m)} & \dots & x_n^{(m)} \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_n \end{pmatrix} \\ &= \begin{pmatrix} 1\theta_0 + x_1^{(1)}\theta_1 + x_2^{(1)}\theta_2 + x_3^{(1)}\theta_3 + \dots + x_n^{(1)}\theta_n \\ 1\theta_0 + x_1^{(2)}\theta_1 + x_2^{(2)}\theta_2 + x_3^{(2)}\theta_3 + \dots + x_n^{(2)}\theta_n \\ 1\theta_0 + x_1^{(3)}\theta_1 + x_2^{(3)}\theta_2 + x_3^{(3)}\theta_3 + \dots + x_n^{(3)}\theta_n \\ \vdots \\ 1\theta_0 + x_1^{(m)}\theta_1 + x_2^{(m)}\theta_2 + x_3^{(m)}\theta_3 + \dots + x_n^{(m)}\theta_n \end{pmatrix}_{m*1} \end{aligned} \tag{6}$$

4. $\vec{y}$

$$\vec{y} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \\ \vdots \\ y^{(m)} \end{pmatrix}_{m*1} \tag{7}$$

1.2.2 Cost Function

1. 数值计算形式:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right]^2 \tag{8}$$

2. 矩阵计算形式:

$$J(\theta) = \frac{1}{2m} \left[h_{\vec{\theta}}(\vec{x}) - \vec{y} \right]^T \left[h_{\vec{\theta}}(\vec{x}) - \vec{y} \right] \tag{9}$$

1.2.3 偏导数$\frac{\partial J(\theta)}{\partial \theta_j}$计算

1. 数值计算形式

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)} \tag{10}$$

2. 矩阵计算形式

$$\nabla J(\theta) = \frac{1}{m} X^T \left[h_{\vec{\theta}}(\vec{x}) - \vec{y} \right] \tag{11}$$

1.2.4 梯度下降迭代方式

1. 数值计算形式

$$
\begin{aligned}
\theta_j &:= \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j} \\
&:= \theta_j - \alpha \frac{1}{m} \left[h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)}
\end{aligned} \tag{12}
$$

2. 矩阵计算形式

$$
\begin{aligned}
\theta &:= \theta - \alpha \nabla J(\theta) \\
&:= \theta - \alpha \frac{1}{m} X^T \left[h_{\vec{\theta}}(\vec{x}) - \vec{y} \right]
\end{aligned} \tag{13}
$$

1.3 Feature Normalization

$$
x_i = \frac{x_i - \mu}{\sigma} \tag{14}
$$

或

$$
x_i = \frac{x_i - \mu}{max - min} \tag{15}
$$

1.4 公式法求解（Normal Equation）

$$
\theta = (X^T X)^{-1} X^T y \tag{16}
$$

2 逻辑回归(Logistic Regression)

2.1 当只有2个类别时，使用1个分类器

2.1.1 预测函数

1. 数值形式

$$h_\theta(x) = \frac{1}{1+e^{\theta^T x}} \tag{17}$$

2. 矩阵形式

$$h_\theta(X) = \frac{1}{1+e^{X\theta}} \tag{18}$$

2.1.2 Cost Function

1. 数值形式

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[-y^{(i)}logh_\theta(x^{(i)}) - (1-y^{(i)})log(1-h_\theta(x^{(i)}))\right] \tag{19}$$

2. 矩阵形式

$$J(\theta) = \frac{1}{m}\left[-y^T \log h_\theta(x) - (1-y^T)\log\left(1-h_\theta(x)\right]\right. \tag{20}$$

2.1.3 梯度下降

1. 数值形式

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left[h_\theta(x^{(i)}) - y^{(i)}\right]x_j^{(i)} \tag{21}$$

迭代方式:

$$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}\left[h_\theta(x^{(i)}) - y^{(i)}\right]x_j^{(i)} \tag{22}$$

2. 矩阵形式

$$\nabla J(\theta) = \frac{1}{m}X^T\left[h_\theta(x) - y\right] \tag{23}$$

迭代方式:

$$\theta := \theta - \alpha\frac{1}{m}X^T\left[h_\theta(x) - y\right] \tag{24}$$

2.2 当只有k个类别时，使用k个分类器