机器学习导论作业二

2020 年 10 月 23 日

181860152 周宇翔

1

(1)

假设对于输入 $\boldsymbol{x_i}=\{x_{i1},x_{i2},...,x_{iK}\}$, 输出为 $\boldsymbol{y_i}=\{y_{i1},y_{i2},...,y_{iL}\}$
令 $\hat{\boldsymbol{x_i}}=(\boldsymbol{x_i},1),i=1,...,m$
$\hat{\boldsymbol{\omega_i}}=(\boldsymbol{\omega_i},b)=(\omega_{i1},\omega_{i2},...,\omega_{iK},b),i=1,...,L$
对于第 i 个标签属性 y_i 的对数似然函数为 $l(\hat{\boldsymbol{\omega_i}})=\sum_{j=1}^{m}lnp(y_{ji}|\boldsymbol{x_j};\hat{\boldsymbol{\omega_i}})$
也即 $l(\hat{\boldsymbol{\omega_i}})=\sum_{j=1}^{m}ln(y_{ji}p_1(\hat{\boldsymbol{x_j}};\hat{\boldsymbol{\omega_i}})+(1-y_{ji})p_0(\hat{\boldsymbol{x_j}};\hat{\boldsymbol{\omega_i}}))$

(2)

最大化上式也即最小化
$r(\hat{\boldsymbol{\omega_i}})=\sum_{j=1}^{m}(-y_{ji}\hat{\boldsymbol{\omega_i}}^T\hat{\boldsymbol{x_j}}+ln(1+e^{\hat{\boldsymbol{\omega_i}}^T\hat{\boldsymbol{x_j}}}))$
这里令$\boldsymbol{\beta}$=$\hat{\boldsymbol{\omega_i}}$
$\nabla r=<\frac{\partial \boldsymbol{r}}{\boldsymbol{\omega_{i1}}},\frac{\partial \boldsymbol{r}}{\boldsymbol{\omega_{i2}}},...,\frac{\partial \boldsymbol{r}}{\boldsymbol{\omega_{iK}}}>$
$=<\sum_{j=1}^{m}-y_{ji}x_{j1}+\frac{x_{j1}e^{\boldsymbol{\beta}^T\hat{\boldsymbol{x_j}}}}{1+e^{\boldsymbol{\beta}^T\hat{\boldsymbol{x_j}}}},....,\sum_{j=1}^{m}-y_{ji}x_{jK}+\frac{x_{jK}e^{\boldsymbol{\beta}^T\hat{\boldsymbol{x_j}}}}{1+e^{\boldsymbol{\beta}^T\hat{\boldsymbol{x_j}}}}>$
可以任取起点, 比如让 $\boldsymbol{\beta_0}=\{0,0,...,0\}$, 取学习率 $\alpha=0.1$
迭代过程中的参数变化为 $\boldsymbol{\beta_{i+1}}=\boldsymbol{\beta_i}-\alpha\nabla(r)$
通过此来完成梯度下降法

2

(3)

我的模型采用了梯度下降法和 OVR 多分类模型, 需要调试的参数主要有学习率 a, 初始系数 $beta$ 以及学习的轮数

在这里采用的梯度下降法非常简单, 即每一轮学习计算当前梯度:$\nabla f = x^T x\beta - x^T y$, 再用系数减去当前梯度 * 学习率. 这里我用的 Sigmoid 函数不是书上的对数几率函数, 因为会产生溢出问题, 而是采用了 logistic function, 它也是一种 S 型曲线函数

训练出 26 个分类器后, 需要采用 OVR 方法对测试集上的数据进行分类, 由于某个数据被多个分类器判为正例或者被所有分类器判为负例的情况都经常发生, 这里不能简单地根据 $logistic\_function(x) > 0.5$ 进行判断, 所以我们选取所有分类器预测出的 $y_i = logistic\_function(x)$ 中最大的 y_i 对应的 i 作为我们最终分类. 这一改进让我的模型预测正确率大幅提升

最终参数:

学习率:0.0000066 初始系数:np.ones((1,17))*0.44 学习轮数:1000

最终性能:

Performance Metric	Value(%)
accuracy	70.42
micro Precision	70.42
micro Recall	70.42
micro F_1	70.42
macro Precision	70.47
macro Recall	70.51
macro F_1	70.49