**机器学习 第一次仿真作业**

1. **题目要求**

用TensorFlow构建一个 SVM分类器，将鸢尾花分为山鸢尾花（Setosa）和非山鸢尾花（non-Setosa）两类 。并利用Tensorboard功能给出计算图的结构，以及一些必要的参数变化曲 ，以及一些必要的参数变化曲线，最后给出测试样本的分类准确率 。

1. **设计SVM分类器**

数据集给出的鸢尾花特征有4个，虽然难以直接可视化，但通过特征两两之间做散点图，不难发现各类鸢尾花的各特征之间有较明显的差别，易于用线性超平面分离，于是选取线性SVM作为分类模型。

把数据集中的四个特征（Sepal length/Sepal width/Petal length/Petal width）直接作为输入向量,输出标签分别表示该样本是否为鸢尾花。SVM分类器的模型分表示为

其中**w**和*b*分别表示各各特征的权重和线性模型的偏置，为待优化的变量。根据SVM线性模型的知识，目标函数可以表示为

其中上式的第二项大于零。求**w**和*b*的最优解的过程可表示为

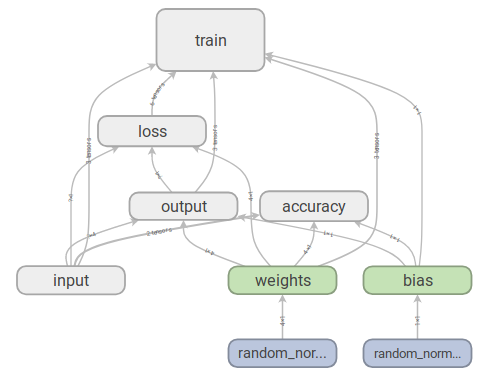
在Tensorflow的框架下，可以利用批量随机梯度下降算法对上式进行求解。另外把数据集分成训练集（80%）和测试集（20%）。经过抽取部分样本的交叉验证，选取模型的主要参数如下表所示。

**表1 训练SVM线性模型的参数**

|  |  |  |
| --- | --- | --- |
| 批量大小BATCH\_SIZE | 迭代次数ITER\_NUM | 学习率LEARNING\_RATE |
| 100 | 10000 | 0.01 |

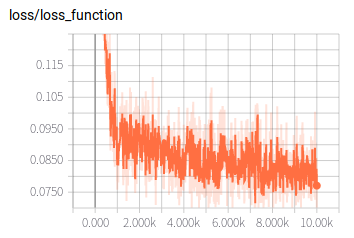
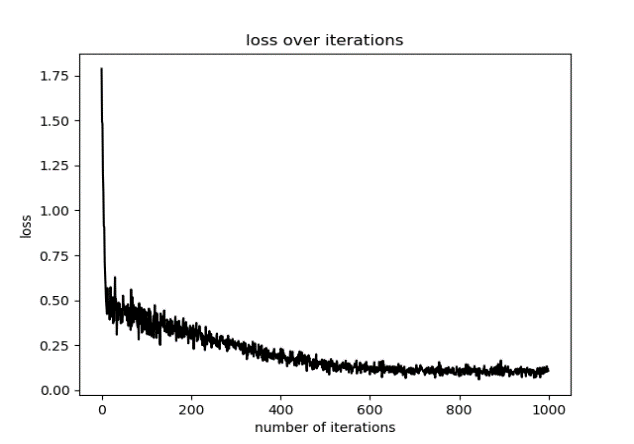
1. **模型结果**

利用Tensorboard对结果可视化。模型的计算图结构如下。



**图1 训练SVM线性模型的计算图结构**

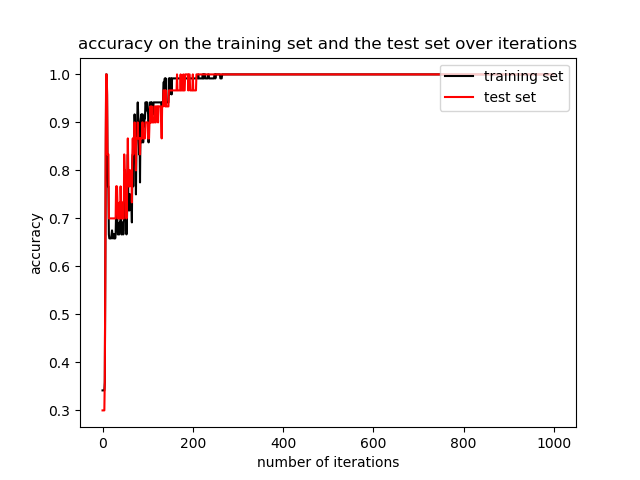
训练过程中目标函数的变化如下，其中(a)利用python工具包绘制，反映了最初1000次迭代的目标函数变化，(b)由Tensorboard给出，反映了目标函数在整个迭代过程中的变化。



1. (b)

**图2 训练过程中的目标函数变化**

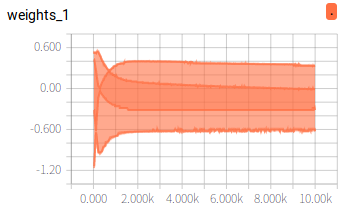
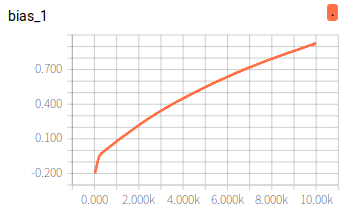
在训练模型的同时，利用测试集对模型的性能进行评估。迭代过程中模型在测试集上的分类准确率如下图所示。(a)(b)两图的关系同上。



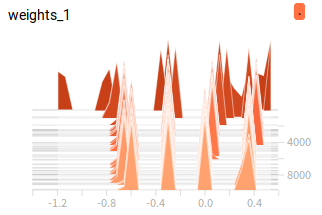
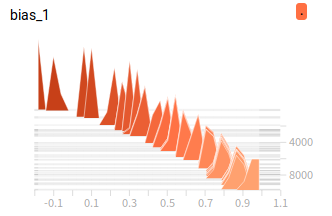
1. (b)

**图3 训练过程中模型在训练集上的分类准确率变化**

另外Tensorboard还给出了**w**和*b*两个变量在训练过程中的变化如下。其中(a)(c)分别是*b* 在训练过程中的直方图和分布图，(b)(d) 分别是**w** 在训练过程中的直方图和分布图。可以看出**w**和*b*在整个训练过程中保持较均匀的变化，没有明显的异常。



1. (b)



(c) (d)

**图4 训练过程中待优化变量w和***b***的变化**

经过迭代的训练， SVM线性分类器在测试集上的准确率可达到100%。