试验和问题

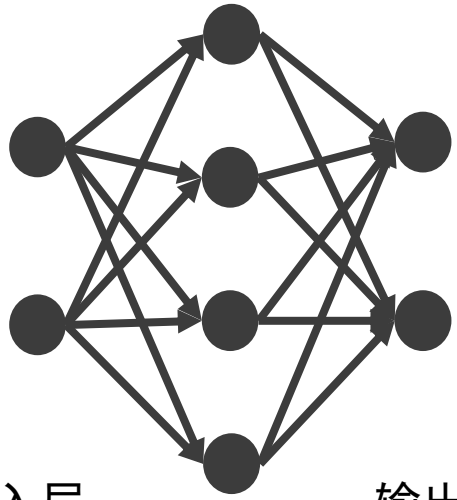
实验环境

TensorFlow1.0.3　Anaconda4.2.0　Python3.5.2　Windows7

实验目的

验证在二分类模型训练完毕后，依据当前分类模型的Uncertainty是否能够指导生成模型生成的数据位于分类边界附近。

生成模型



输入层　　输出层

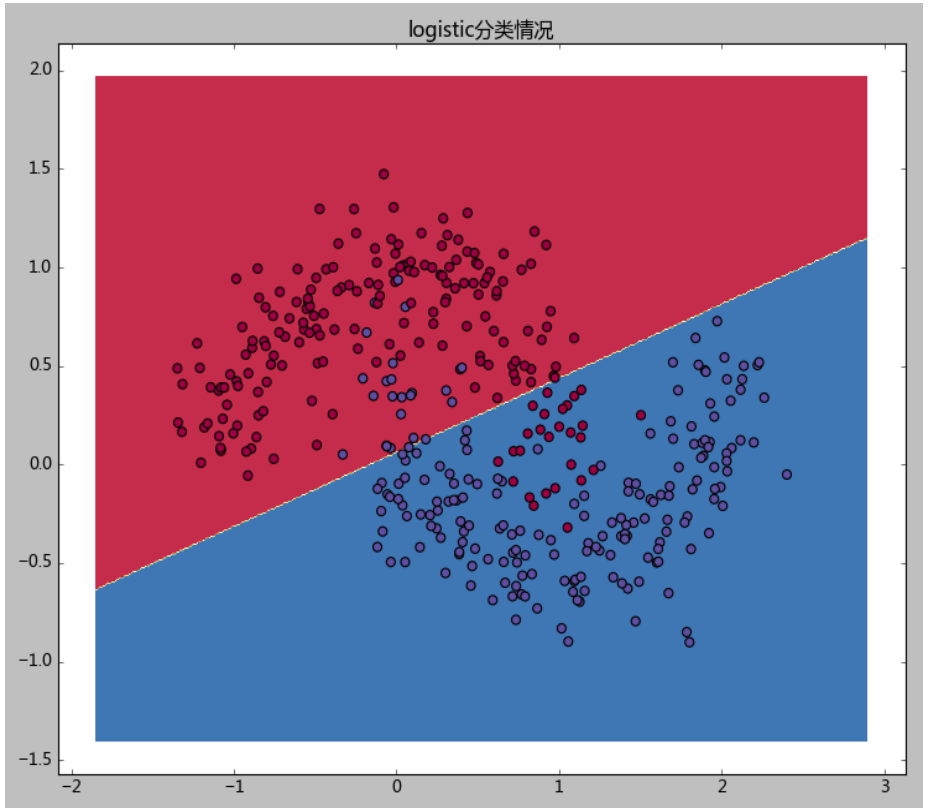
tanh

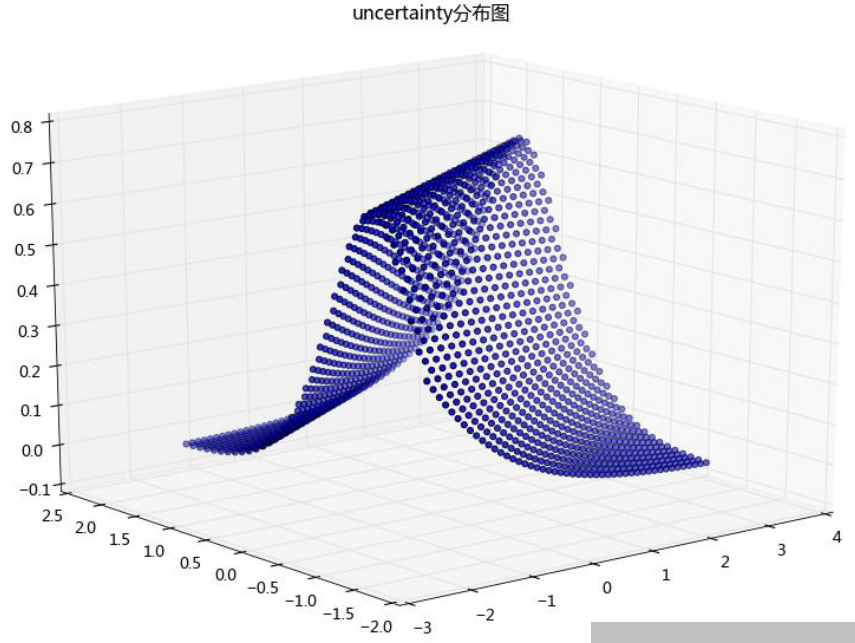
分类模型

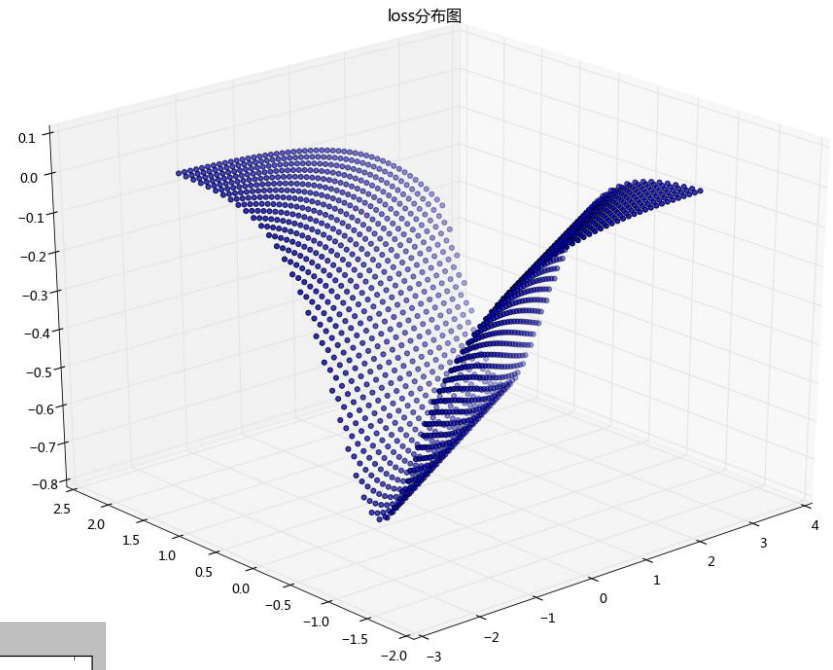
$$p_1 = \frac{\exp(w \cdot x)}{1 + \exp(w \cdot x)}$$

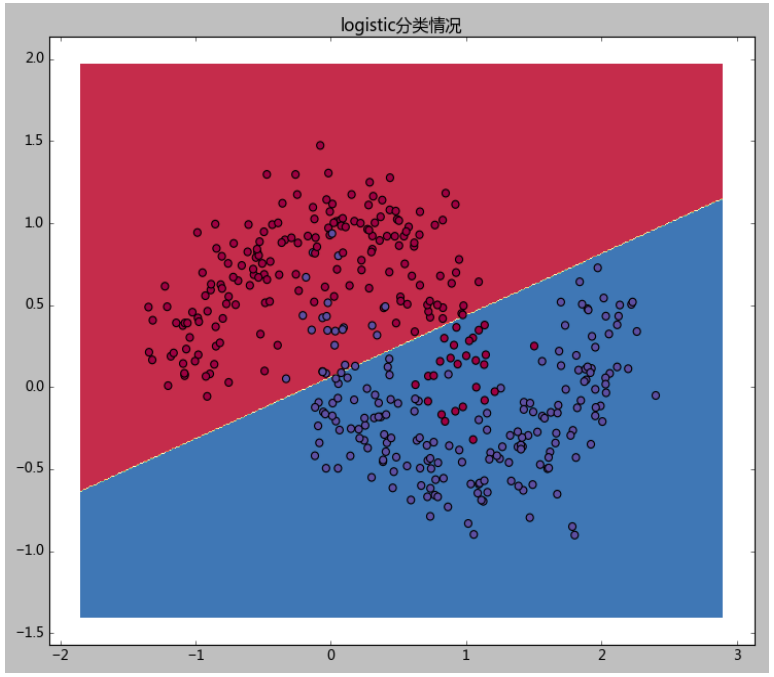
$$p_2 = \frac{1}{1 + \exp(w \cdot x)}$$

$$w \cdot x = w_1 \cdot x_1 + w_2 \cdot x_2$$

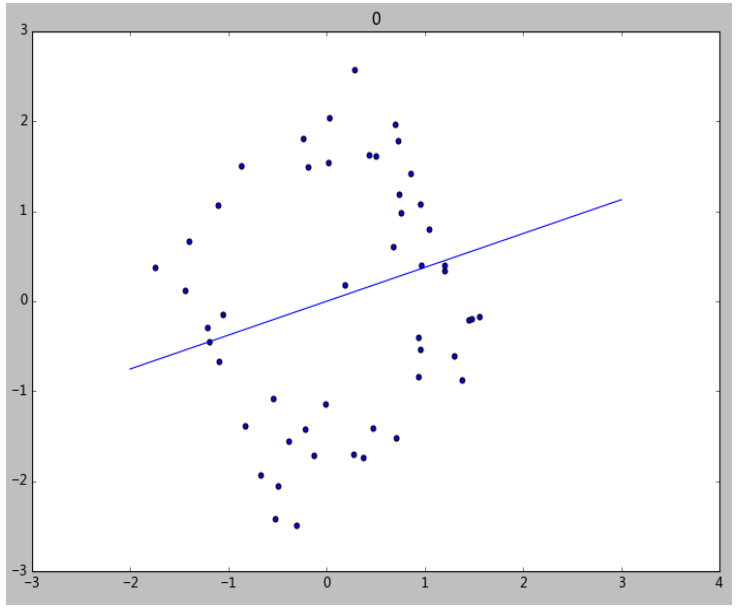


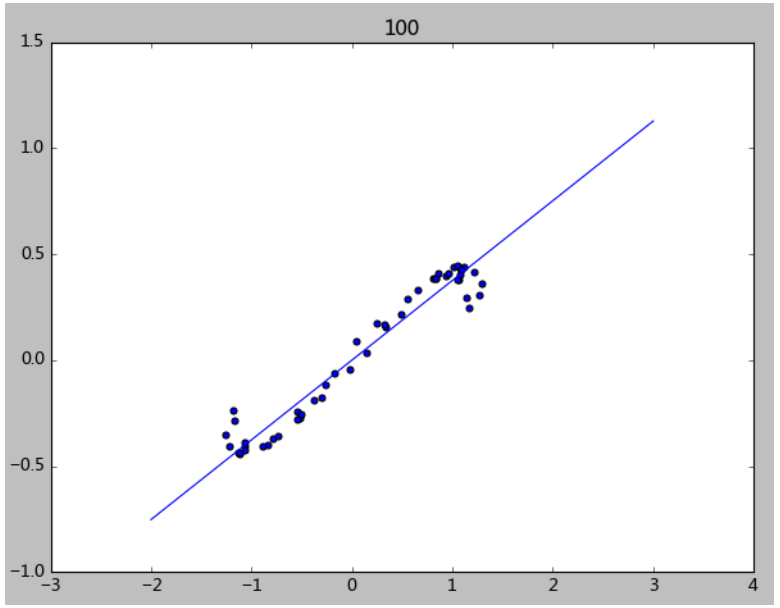


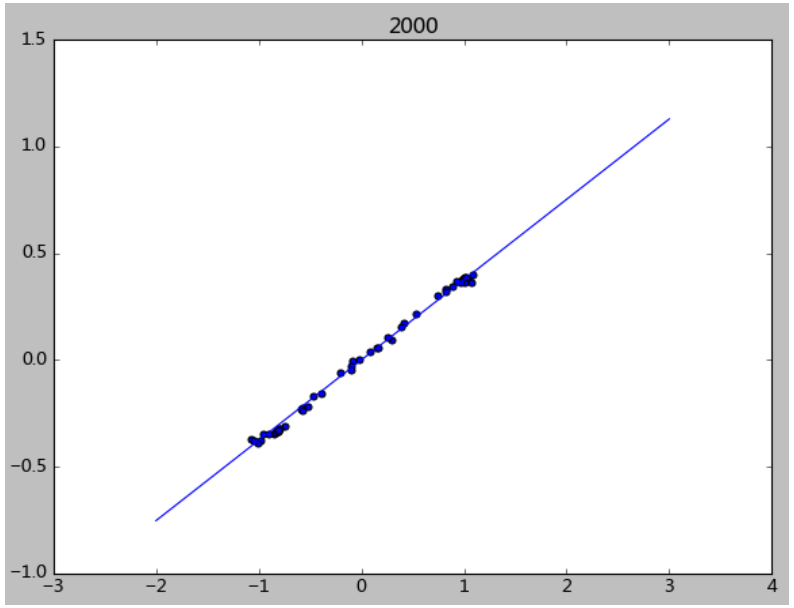


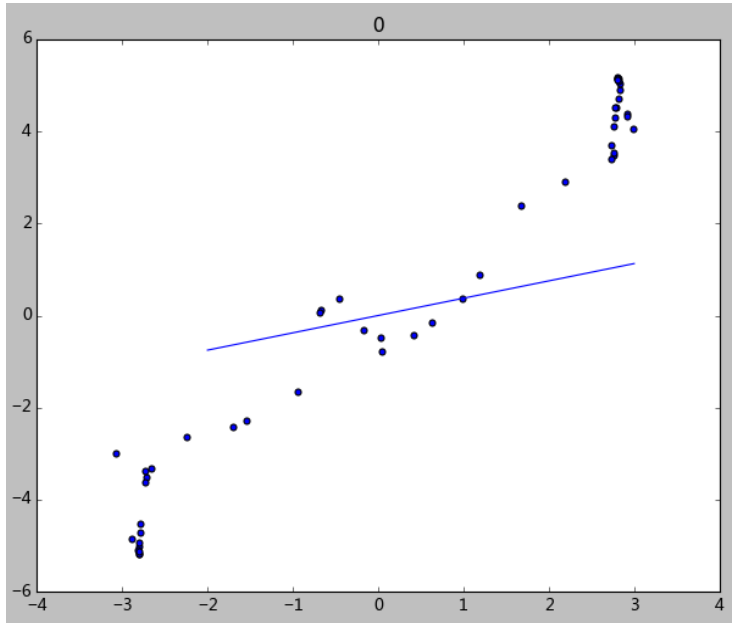


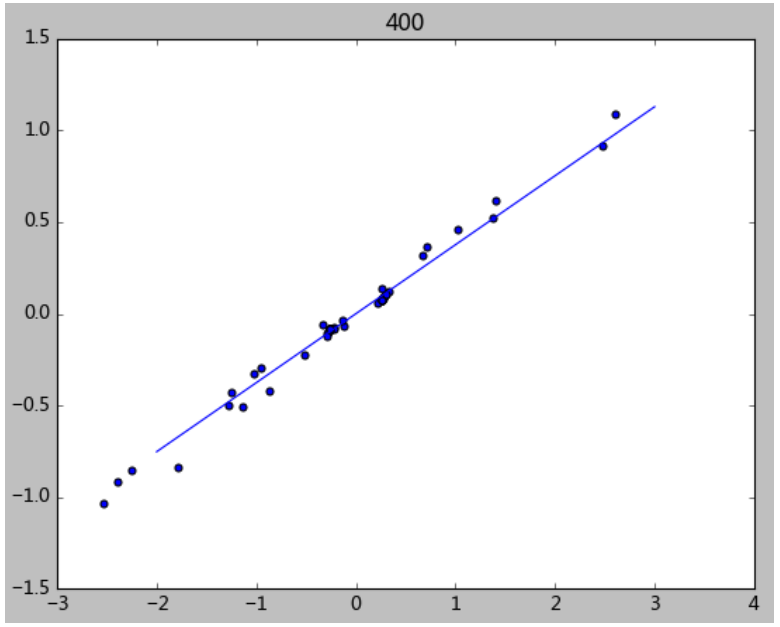
$$x_H^* = \operatorname*{argmax}_x - \sum_i P_\theta(y_i|x) \log P_\theta(y_i|x)$$

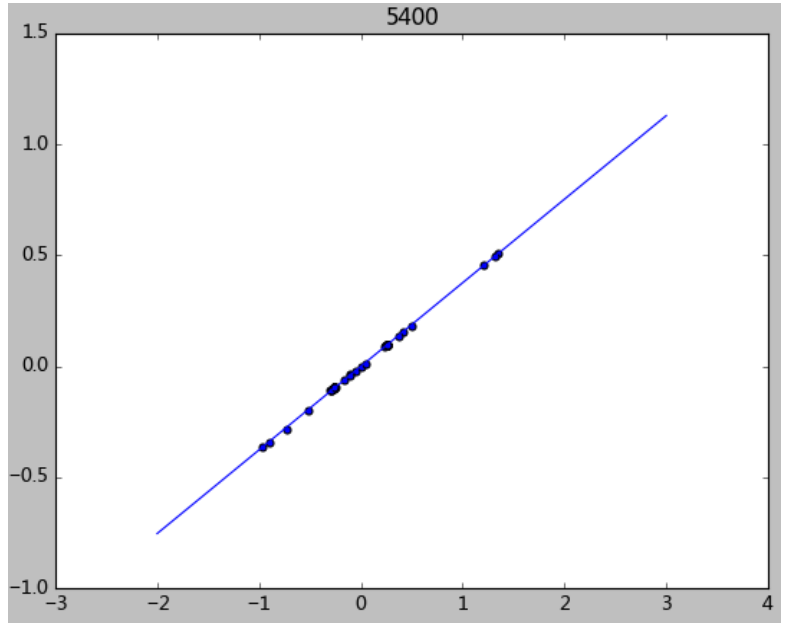

0

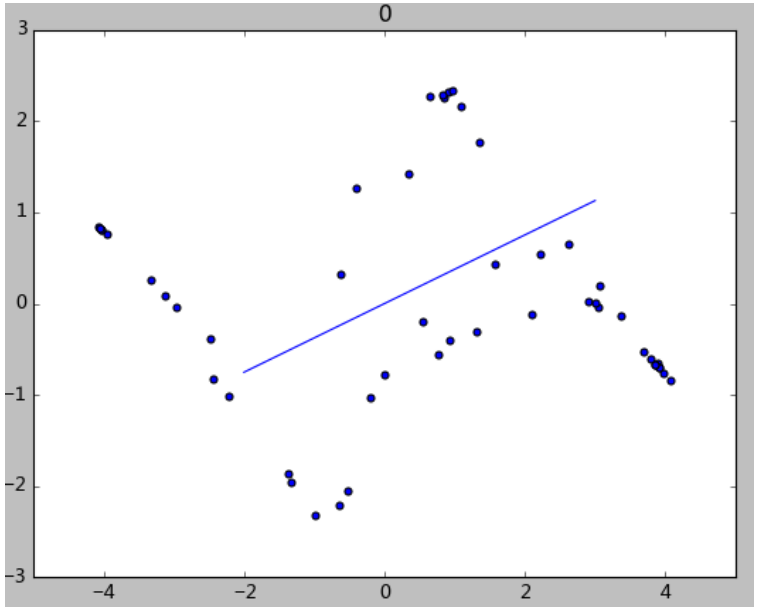

100

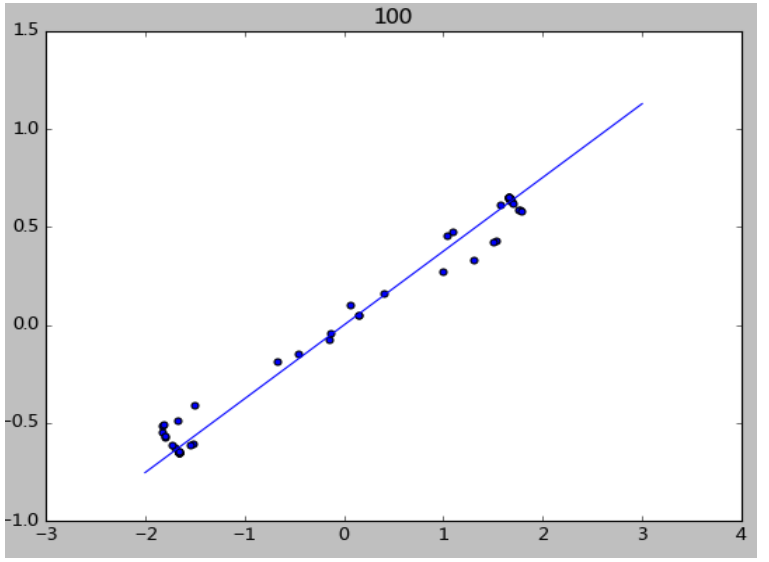

2000

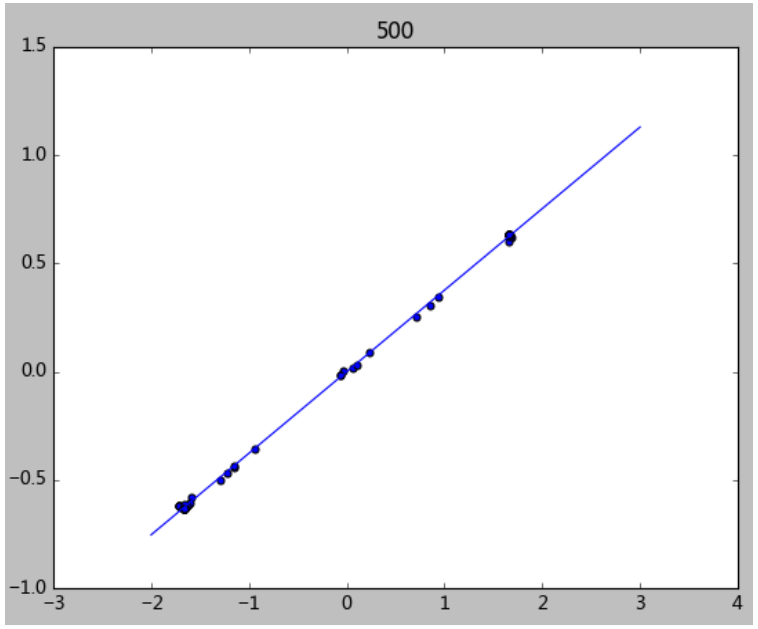

0


400


5400







$$x_M^* = \underset{x}{\operatorname{argmin}}\, P_\theta(\hat{y}_1|x) - P_\theta(\hat{y}_2|x)$$

$$loss = (p_1 - p_2)^2$$

问题

引入和数据点密度分布相关信息

覆盖数据特征空间范围

均匀分布在现有数据集分类界面

Uncertainty

多分类问题不再适用

$$x_H^* = \operatorname*{argmax}_x - \sum_i P_\theta(y_i|x) \log P_\theta(y_i|x)$$

$$x_M^* = \operatorname*{argmin}_x P_\theta(\hat{y}_1|x) - P_\theta(\hat{y}_2|x)$$

模型选择

Relevance Vector Machine

1.寻找合适的大规模数据预处理方式（聚类）

2.实现RVM多分类应用

3.Loss设计选择

4.基于图像数据集的LSH实现