

Homework2

刘喆骐 2020013163 探微化01

1

(1)

模型可能会发生过拟合。

区别：验证集用于进一步确定模型中的超参数，即调整模型的结构和参数，以提高模型的泛化能力；而测试集则是用来评估模型的泛化能力，即模型对未知数据的预测能力。

(2)

将数据集分成K份，每次拿出一份作为验证集，剩下k-1份作为训练集。重复K次，每次选择不同的验证集，最后将K次的结果取平均。

(3)

L1正则化可以产生稀疏权重矩阵，即产生一个稀疏模型，将一些权重减为0，可以用于特征选择。

弹性网络正则化是一种结合了L1和L2正则化的方法，它同时惩罚权重系数的绝对值和平方的值，可以用于特征选择和防止过拟合。弹性网络的优化目标是：
$$\min_w \frac{1}{2n} \|Xw - y\|_2^2 + \alpha \rho \|w\|_1 + 2\alpha(1 - \rho) \|w\|_2^2$$

(4)

引入核函数的动机：在低维空间中，数据可能是线性不可分的，但是在高维空间中，数据可能是线性可分的。因此，我们可以通过核函数将数据从低维空间映射到高维空间来解决这个问题。

(5)

1. 自助采样法: 对于每棵树, 从原始数据集中有放回地随机抽取相同大小的样本。
2. 随机选择特征: 在每个节点进行分裂时, 从所有特征中随机选择一定数量的特征作为候选特征。

2

(1)

增加训练数据量，减少模型的复杂度。

正则化.

Dropout: 在训练过程中随机删除一些神经元。

Early stopping: 在训练过程中, 当验证集上的误差不再下降时停止训练。

(2)

平移不变性和局部性。局部性是指图像中的每个区域只与其周围的一小部分区域相关, 而平移不变性是指图像中的特征在不同位置上具有相同的意义。

(3)

LeNet 共有卷积层 (C1), 池化层 (S2), 卷积层 (C3), 池化层 (S4), 全连接层 (F5), 全连接层(F6), 输出层 (F7) 共 7 层, 其中包含可学习参数的层为:C1、C3、F5、F6 和 F7。

不考虑偏置参数, 这些层的可学习参数量分别为:

$$C1 : 6 \times 5 \times 5 = 150$$

$$C3 : 16 \times 5 \times 5 = 400$$

$$F5 : 120 \times (16 \times 5 \times 5) = 48000$$

$$F6 : 84 \times 120 = 10080$$

$$F7 : 10 \times 84 = 840$$

(4)

ResNet在模型架构上提出了残差学习的思想, 即在网络中增加了跨层连接, 使得网络可以更深, 同时避免了梯度消失的问题。

(5)

gradient clipping: 梯度大于某阈值时进行裁剪, 使之不大于该阈值。

Variational Dropout: 在每个时间步丢弃相同的网络单元, 随机丢弃输入、输出和循环连接。



清华大学探微书院
Tanwei College, Tsinghua University

作业纸

班级: 姓名: 学号: 科目: 第 页

11

根蒂: $H(D) = 0.998$

$$H(D_1) = -\left(\frac{5}{8} \lg \frac{5}{8} + \frac{3}{8} \lg \frac{3}{8}\right) = 0.955$$

$$H(D_2) = -\left(\frac{3}{7} \lg \frac{3}{7} + \frac{4}{7} \lg \frac{4}{7}\right) = 0.985$$

$$H(D_3) = 0$$

$$I_A = 0.143$$

高皮声: $H(D) = 0.998$

$$H(D_1) = -\left(\frac{6}{10} \lg \frac{6}{10} + \frac{4}{10} \lg \frac{4}{10}\right) = 0.971$$

$$H(D_2) = -\left(\frac{2}{5} \lg \frac{2}{5} + \frac{3}{5} \lg \frac{3}{5}\right) = 0.971$$

$$H(D_3) = 0$$

$$I_A = 0.141$$

纹理: $H(D) = 0.998$

$$H(D_1) = -\left(\frac{2}{3} \lg \frac{2}{3} + \frac{1}{3} \lg \frac{1}{3}\right) = 0.764$$

$$H(D_2) = -\left(\frac{1}{5} \lg \frac{1}{5} + \frac{4}{5} \lg \frac{4}{5}\right) = 0.722$$

$$H(D_3) = 0$$

$$I_A = 0.381$$

声部: $H(D) = 0.998$

$$H(D_1) = -\left(\frac{5}{7} \lg \frac{5}{7} + \frac{2}{7} \lg \frac{2}{7}\right) = 0.963$$

$$H(D_2) = -\left(\frac{3}{6} \lg \frac{3}{6} + \frac{3}{6} \lg \frac{3}{6}\right) = 1$$

$$H(D_3) = 0$$

$$I_A = 0.290$$

情感: $H(D) = 0.998$

$$H(D_1) = -\left(\frac{6}{12} \lg \frac{6}{12} + \frac{6}{12} \lg \frac{6}{12}\right) = 1$$

$$H(D_2) = -\left(\frac{2}{5} \lg \frac{2}{5} + \frac{3}{5} \lg \frac{3}{5}\right) = 0.971$$

~~1.143~~

$$I_A = 6.53 \times 10^{-3}$$



清华大学探微书院



作业纸

班级:

姓名:

学号:

科目:

第 页

(2) ~~西瓜~~ 纹理与熵最大, 按此纹理分类: 清晰、稍模糊、模糊。

对于稍模糊西瓜, 可以由颜色直接分类: 软糯是西瓜, 硬滑为坏瓜。

对于纹理清晰的西瓜:

$$H(D) = 0.764$$

$$\begin{aligned} \text{色泽: } H(D_1) &= 0.811 & \text{根蒂: } H(D_1) &= 0 & \text{声音: } H(D_1) &= 0.15 \\ H(D_2) &= 0.811 & H(D_2) &= 0.918 & H(D_2) &= 0.50 \\ H(D_3) &= 0 & H(D_3) &= 0 & H(D_3) &= 0 \\ I_G &= 0.043 & I_G &= 0.458 & I_G &= 0.331 \end{aligned}$$

$$\begin{aligned} \text{脐部: } H(D_1) &= 0 & \text{脐部: } H(D_1) &= 0 \\ H(D_2) &= 0.918 & H(D_2) &= 0.918 \\ H(D_3) &= 0 & I_G &= 0.458 \\ I_G &= 0.458 \end{aligned}$$

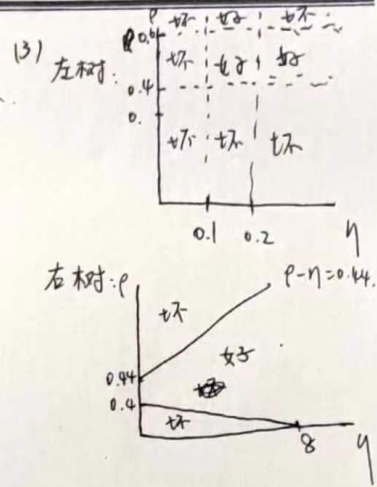
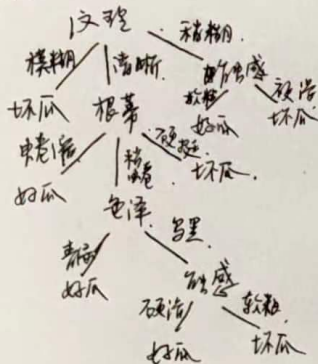
由于根蒂在前, 选择根蒂来分类: 软糯、稍硬、硬挺。

对于稍硬: 2个为西瓜, 1个为坏瓜。

由于色泽和脐部与熵相同, 故用色泽分类。

剩下两个西瓜, 一好一坏, 用色泽分类。

决策树如下:



$$l = -\log\left(\frac{e^{o_k}}{\sum_{j=1}^K e^{o_j}}\right) = -o_k + \log\left(\sum_{j=1}^K e^{o_j}\right)$$

当 $i \neq k$ 时,

$$\frac{\partial l}{\partial o_i} = \frac{\partial \log(\sum_{j=1}^K e^{o_j})}{\partial o_i} = \frac{e^{o_i}}{\sum_{j=1}^K e^{o_j}}$$

当 $i = k$ 时,

$$\frac{\partial l}{\partial o_i} = -1 + \frac{\partial \log(\sum_{j=1}^K e^{o_j})}{\partial o_i} = -1 + \frac{e^{o_i}}{\sum_{j=1}^K e^{o_j}}$$

故

$$\frac{\partial l}{\partial o} = \left(\frac{e^{o_1}}{\sum_{j=1}^K e^{o_j}}, \frac{e^{o_2}}{\sum_{j=1}^K e^{o_j}}, \dots, \frac{e^{o_{k-1}}}{\sum_{j=1}^K e^{o_j}}, \frac{e^{o_k}}{\sum_{j=1}^K e^{o_j}}, \frac{e^{o_{k+1}}}{\sum_{j=1}^K e^{o_j}}, \dots, \frac{e^{o_K}}{\sum_{j=1}^K e^{o_j}} \right)$$

$$\frac{\partial l}{\partial o} = \hat{y} - y$$

(2)

(2.1)

$$\frac{\partial l}{\partial W} = \frac{\partial l}{\partial o} \frac{\partial o}{\partial W}$$

$$o = W^T x, o_i = \sum_{j=1}^D W_{ij} x_j$$

故

$$\frac{\partial o_i}{\partial W_{ij}} = x_j$$

$$\frac{\partial l}{\partial W_{ij}} = \sum_{k=1}^K \frac{\partial l}{\partial o_k} \frac{\partial o_k}{\partial W_{ij}} = (\hat{y}_i - y_i) x_j^T$$

故

$$\frac{\partial l}{\partial W} = x(\hat{y} - y)^T$$

(2.2)

$$\frac{\partial l}{\partial x} = \frac{\partial l}{\partial o} \frac{\partial o}{\partial x}$$

$$\frac{\partial o_i}{\partial x_j} = W_{ij}$$

$$\frac{\partial l}{\partial x_j} = \sum_{k=1}^K \frac{\partial l}{\partial o_k} \frac{\partial o_k}{\partial x_j} = \sum_{k=1}^K (\hat{y}_k - y_k) W_{kj}$$

故

$$\frac{\partial l}{\partial x} = W(\hat{y} - y)$$

(3)

$$\begin{aligned} \frac{\partial h_i}{\partial z_i} &= \frac{\partial \frac{1}{1+e^{-z_i}}}{\partial z_i} = \frac{e_i^z}{(1+e_i^z)^2} = h_i(1-h_i) \\ \frac{\partial l}{\partial z} &= [\sum_{j=1}^H \frac{\partial l}{\partial h_j} \frac{\partial h_j}{\partial z_j}]_{i=1}^H = [h_i(1-h_i)W_2(\hat{y}-y)]_{i=1}^H = h \odot (1-h) \odot W_2(\hat{y}-y) \\ \frac{\partial l}{\partial W_1} &= \frac{\partial l}{\partial z} \frac{\partial z}{\partial W_1} \\ z_i &= \sum_{j=1}^D W_{1ij}x_j \end{aligned}$$

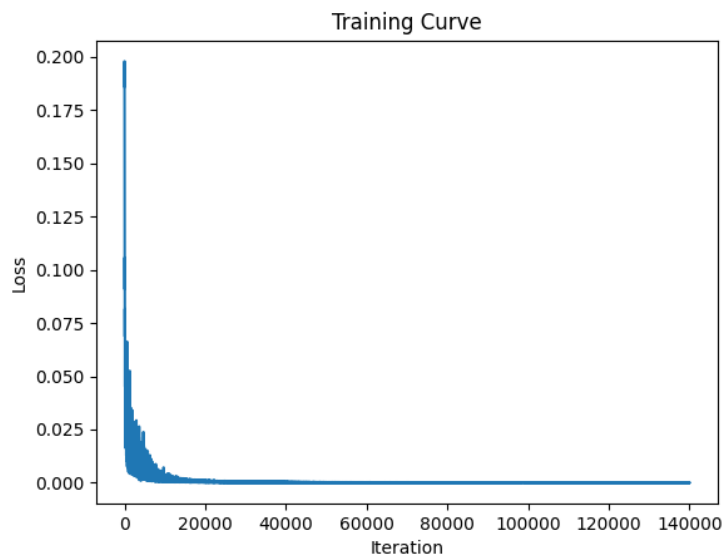
故

$$\begin{aligned} \frac{\partial z_i}{\partial W_{1ij}} &= x_j \\ \frac{\partial l}{\partial W_{1ij}} &= \sum_{k=1}^H \frac{\partial l}{\partial z_k} \frac{\partial z_k}{\partial W_{1ij}} = h_i(1-h_i)(W_2(\hat{y}-y))_i x_i \end{aligned}$$

故

$$\frac{\partial l}{\partial W_1} = x[h \odot (1-h) \odot W_2(\hat{y}-y)]^T$$

(4)



loss 图

最佳选择：hidden_dim=75, lr=0.005, batch_size=64, epochs=1000

结果：训练集准确率1，验证集准确率0.959，测试集准确率0.9502