

第一次作业

自硕 21 崔晏菲 2017012326

1. 人脸识别，是基于人的脸部特征信息进行身份识别的一种生物识别技术。用摄像机或摄像头采集含有人脸的图像或视频流，并自动在图像中检测和跟踪人脸，进而对检测到的人脸进行脸部识别的一系列相关技术，通常也叫做人像识别、面部识别。其原理大概是先提取人脸的一些特定的模式，也叫特征，接着将这些特征与数据库中的其他人脸进行比较，最后得出识别结果。人脸识别有很多种方法，早起人们手动进行特征的提取，使用 PCA、Isomap 等方法提取特征，并用 SVM 进行分类。而现在，人工神经网络可以自动进行特征的提取。这之所以是一个机器学习技术，是因为这并不是人类手动的将这些特征“告诉”机器，而是让机器自动地将对应的特征提取出来并进行比较，在只依赖原始数据的情况下，机器自己“学会”了人脸识别。

2. 易知， d 维空间中的线性决策面维

$$\begin{aligned} f(\mathbf{x}) &= w_0 + \sum_{i=1}^d w_i x_i \\ &= \mathbf{w}^T \cdot \mathbf{x} + w_0 = 0 \end{aligned}$$

其中, $\mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$

易知，对于决策面上的任意两点 $\mathbf{x}_1, \mathbf{x}_2$ 都有

$$f(\mathbf{x}_1) = f(\mathbf{x}_2) = 0$$

即

$$\mathbf{w}^T(\mathbf{x}_2 - \mathbf{x}_1) = 0$$

故 \mathbf{w} 是决策面的法向量。那么对于空间中任意一点 \mathbf{x}_0 ，我们记它与决策面的距离为 r ，设 \mathbf{x}_p 是 \mathbf{x}_0 在决策面方向上的投影，则有

$$\mathbf{x}_0 = \mathbf{x}_p + r \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

则

$$\begin{aligned} f(\mathbf{x}_0) &= f\left(\mathbf{x}_p + r \frac{\mathbf{w}}{\|\mathbf{w}\|}\right) \\ &= f(\mathbf{x}_p) + f\left(r \frac{\mathbf{w}}{\|\mathbf{w}\|}\right) - w_0 \\ &= r \frac{\mathbf{w}^T \mathbf{w}}{\|\mathbf{w}\|} = r \|\mathbf{w}\| \end{aligned}$$

故

$$r = \frac{f(\mathbf{x}_0)}{\|\mathbf{w}\|}$$

特别地，原点到决策面的距离为

$$\frac{w_0}{\|\mathbf{w}\|}$$

3. 记事件 T 为“得病”， F 为“没得病”， Y 为“检测到阳性”， Z 为“检测到阴性”。

(1) 在这个城市里患病的先验概率为 $P(T) = \frac{100}{1000000} = 0.0001$

对于产品 A 来说，若被检测为阳性，那么真实患病的可能性为

$$\begin{aligned} P(T|Y_A) &= \frac{P(Y_A|T) \cdot P(T)}{P(Y_A|T) \cdot P(T) + P(Y_A|F) \cdot P(F)} \\ &= \frac{0.7 \times 0.0001}{0.7 \times 0.0001 + 0 \times (1 - 0.0001)} = 1 \end{aligned}$$

用产品 A 检测到阴性，那么真实患病的可能性为

$$P(T|Z_A) = \frac{P(Z_A|T) \cdot P(T)}{P(Z_A|T) \cdot P(T) + P(Z_A|F) \cdot P(F)}$$

$$= \frac{0.3 \times 0.0001}{0.3 \times 0.0001 + 1 \times (1 - 0.0001)} = 3.000210014701029 \times 10^{-5}$$

对于产品 B 来说，若被检测为阳性，那么真实患病的可能性为

$$P(T|Y_B) = \frac{P(Y_B|T) \cdot P(T)}{P(Y_B|T) \cdot P(T) + P(Y_B|F) \cdot P(F)} = 0.008920606601248876$$

用产品 B 检测到阴性，那么真实患病的可能性为

$$P(T|Z_B) = \frac{P(Z_B|T) \cdot P(T)}{P(Z_B|T) \cdot P(T) + P(Z_B|F) \cdot P(F)} = 1.0101918253257108 \times 10^{-5}$$

对于产品 C 来说，若被检测为阳性，那么真实患病的可能性为

$$P(T|Y_C) = 0.00032992411745298575$$

用产品 C 检测到阴性，那么真实患病的可能性为

$$P(T|Z_C) = 1.4287122587797954 \times 10^{-6}$$

故我建议用产品 A

(2) 产品 A 可以有助于维持正常的社交生活，但是假阴性率较大，不利于疫情防控。

不论产品 B 检测到阳性还是阴性，患病的概率都比产品 C 高。而产品 C 的假阴性率较低，同时假阳性率很高，可以有效防控疫情，但是不利于维持正常社交生活。