Project Part-1

After feature extraction – 6000×2 (Class 0)

Training size – 12000 × 2 ≈ 784

Test data size = 2000 × 2

① Estimation parameters – np.cov

One class: $\mu = [\mu_1, \mu_2]$, $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$

Class 1 – $\mu = [\mu_1, \mu_2]$, Σ

Class 0 – $\mu = [\mu_1, \mu_2]$, Σ

Hint: The features are independent. $\begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}$

Class 0 – Avg, standard deviation

Avg → 1-D Gaussian; Std → 1-D Gaussian

Class 1 – Avg → 1-D; Std → 1-D Gaussian

784 → 1

② Naive Bayes –

$P(Y|X) \sim P(X|Y) \cdot P(Y)$ (From the data)

$P(X_1|Y) \cdot P(X_2|Y) \cdot P(Y)$

$P(Y=0|x) > P(Y=1|x) \rightarrow$ Class 0

otherwise Class = 1

③ Logistic Regression – weights w $P(Y|X)$

Step 1: Initialize w randomly. $[w_0, w_1, w_2]$

Step 2: $z = \sigma(w^T x)$ $\sigma(t) = \frac{1}{1+e^{-t}}$

$w_0 + w_1 x_1 + w_2 x_2$

Step 3: Calculate the cost function.

Log-likelihood func^n –

$$L(w) = \sum_{i=1}^{m} y^{(i)} \log z^{(i)} + (1-y^{(i)}) \log(1-z^{(i)})$$

$\leftarrow e^{-8}$

Step 4: Gradient-ascent np.random

$$w_{t+1}^j = w_t^j + \eta \frac{\partial L(w)}{\partial w^j}$$

iteration

$$\frac{\partial L(w)}{\partial w^j} = \frac{\partial L(w)}{\partial z} \cdot \frac{\partial z}{\partial p} \cdot \frac{\partial p}{\partial w^j} \Rightarrow \text{Using chain rule.}$$

$$\frac{\partial L(w)}{\partial z} = \frac{y}{z} - \frac{(1-y)}{1-z} \quad -(1)$$

$$\frac{\partial z}{\partial p} = z(1-z) \quad -(2)$$

$$\frac{\partial p}{\partial w^j} = x^j \quad -(3)$$

$$\frac{\partial L(w)}{\partial w^j} = \frac{y(1-z) - z(1-y)}{z(1-z)} x^j$$

$$= (y-z)\, x^j$$

Prediction $z = \sigma(w^T x)$

$z > 0.5 \rightarrow$ Class 1

otherwise $\rightarrow$ Class 0

Covariance matrix – x_1 x_2

6000 × 2 ⇒ Class 0

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}$$

12000 × 2 (x_1 x_2) – Training data

6000 × 2 Class 0 → μ_1, μ_2, Σ

6000 × 2 Class 1 → μ_1, μ_2, Σ