

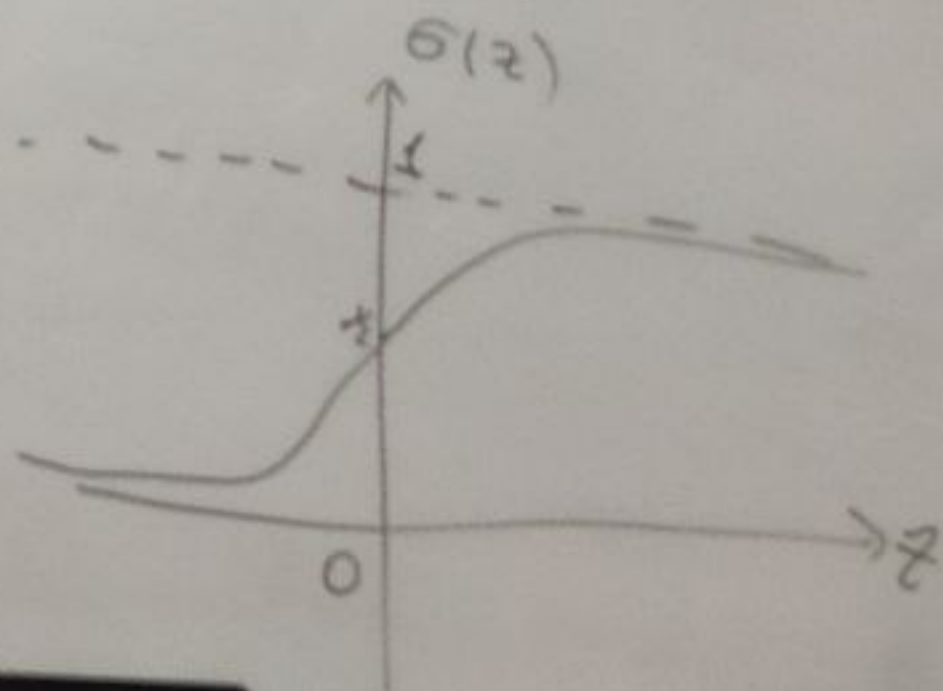
Логистическая регрессия

$y = \{0, 1\}$ - двухклассовая регр.

y_1, \dots, y_n - выборка, где $y_i \sim \text{Bern}(\mu_\theta(x_i))$

$$\mu_\theta(x_i) = \sigma(\theta^T x_i)$$

$$\sigma(z) = \frac{1}{1 + e^{-z}} \text{ - сигмоида}$$



Своо

$\sigma(z)$

(классическая) σ -алгебра
 $\{ \sigma(x) = \frac{1}{2} \} = \{ \sigma^2(x) = 0 \}$ - дискретная σ -алгебра
 σ - фактор \Rightarrow почти everywhere сходимость
 (классическая) σ -алгебра
 ① $\sigma(-\sigma) = 1 - \sigma(\sigma)$
 ② $\sigma^2(\sigma) \cap \sigma(\sigma) = \emptyset$ $\Rightarrow \frac{1}{1-\sigma} = \text{almost everywhere}$
 ③ $\sigma^2(\sigma) = 1 - \sigma(\sigma)$

(классическая) σ -алгебра
 Question 0-2
 $(X_1, Y_1), \dots, (X_n, Y_n)$ - independent
 на Σ σ -алгебра $Y_i \sim \text{Bern}(p(x_i))$



Логистическая регрессия

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{i=1}^n \left[y_i (1 - \sigma(\theta^T x_i)) + (1 - y_i) \cdot (-\sigma(\theta^T x_i)) \right] x_i =$$

$$= \sum_{i=1}^n (y_i - \sigma(\theta^T x_i)) x_i = X^T (Y - S(\theta))$$

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad S(\theta) = \begin{pmatrix} \sigma(\theta^T x_1) \\ \vdots \\ \sigma(\theta^T x_n) \end{pmatrix}$$

Свойства лог. ф.

$$\{ \sigma(\theta^T x) = \frac{1}{2} \}$$

σ — расчёт —

Свойства сиг.

① $\sigma(-z)$

② $\sigma'(z)$

③ $\sigma'(z)$

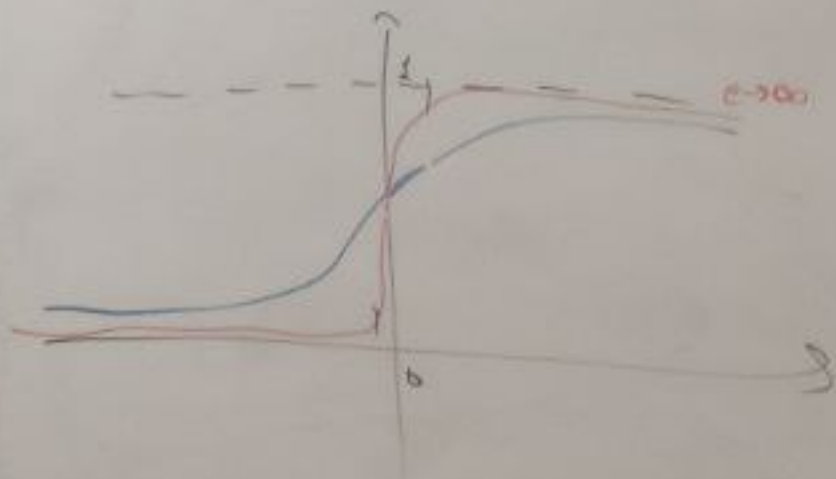
Классификация

Переобучение

- Рассмотрим случаи:
- классы линейно, разделимы
 - среди признаков есть единич.
 - $\exists \theta, \gamma, \epsilon, \{ \theta^T x \geq 0, \gamma - \text{разделение классов} \}$

Логистическая регрессия
 $\forall c > 0 \in \mathbb{R}: \delta c \theta^T x = 0.5 = h \theta^T x = 0.5$ -
 -разделяющая гиперплоскость

$\sigma(\underbrace{c \theta^T x}_{\rightarrow 0}) \xrightarrow{c \rightarrow \infty} 1$, приближается к 1.



Проблемы

• мн. п.н. либо ~ 0
 либо ~ 1

• $\theta^T x \approx 0 \rightarrow$

\Rightarrow сложно классифицировать

Градиентный спуск (n)

$$\theta_{k+1} = \theta_k + \eta \nabla F(\theta_k)$$

$$\theta_{k+1} = \theta_k + \eta X^T ($$



IRIS / Iterative Recursive Least Squares
Mixed Minimization
Normalized Gradient

$F(\theta) = y - X\theta$
 $\theta_{new} = \theta_{old} - \eta \nabla F(\theta)$
 $\nabla F(\theta) = -X^T (y - X\theta)$
 $DF(\theta) = -X^T (y - X\theta)$
 $DF(\theta) = -X^T y + X^T X \theta$
 $DF(\theta) = -X^T y + X^T X \theta$

Normalized Gradient
 $\theta_{new} = \theta_{old} + \eta \frac{\nabla F(\theta)}{\|\nabla F(\theta)\|}$

(Iterative Method (Gauss-Seidel))
 $\theta_{k+1} = \theta_k + \alpha \nabla_{\theta} L(\theta_k)$
 Gauss-Seidel
 $\theta_{k+1} = \theta_k + \alpha \nabla_{\theta} L(\theta_k)$
 Gauss-Seidel (HMM)
 $\theta_k = \begin{bmatrix} y - X\theta_k \\ y - X\theta_k \end{bmatrix}$
 $\theta_k = 2X^T(y - X\theta_k) = 2X^T y - 2X^T X \theta_k$
 $\theta_{k+1} = \theta_k + (X^T y - X^T X \theta_k) =$

Gauss-Seidel
 $\theta_{k+1} = \theta_k + (X^T y - X^T X \theta_k) = (X^T y - X^T X \theta_k) =$
 $\frac{1}{2} (X^T y - X^T X \theta_k) = \text{gauss-seidel}$
 Gauss-Seidel
 $F(\theta) = L_{\text{HMM}}(\theta) = \sum_{i=1}^n \left[y_i \ln(\theta^T x_i) + (1 - y_i) \ln(1 - \theta^T x_i) \right]$
 $\nabla F(\theta) = - \sum_{i=1}^n (y_i - \theta^T x_i) x_i = -X^T (y - X\theta)$



Логистическая регрессия

$$\begin{aligned}
 \text{DD } F(\theta) &= \left(\sum_{i=1}^n \sigma(\theta^T x_i) x_i \right)' = \\
 &= \sum_{i=1}^n \left[x_i^T \sigma(\theta^T x_i) \cdot (1 - \sigma(\theta^T x_i)) x_i \right] \\
 &= X^T V(\theta) X
 \end{aligned}$$

$\begin{matrix} d \times n & n \times n & n \times d \end{matrix}$

$$V(\theta) = \text{diag}(\sigma(\theta^T x_1)(1 - \sigma(\theta^T x_1)) \dots \sigma(\theta^T x_n)(1 - \sigma(\theta^T x_n)))$$

$$\begin{aligned}
 \theta_{k+1} &= \theta_k - (X^T V(\theta_k) X)^{-1} X^T (S(\theta_k) - y) = \\
 &= (X^T V(\theta) X)^{-1} \left[X^T V(\theta) X \theta_k - X^T (S(\theta_k) - y) \right] =
 \end{aligned}$$

IRLS (Iterative Reweighted Least Squares)

$$= (X^T V(\theta_k) X)^{-1} X^T V(\theta_k) z(\theta_k)$$

$$z(\theta_k) = [X\theta_k - V(\theta_k)^{-1} (S(\theta_k) - y)]$$