Лекция 9. Линейные модели. Часть 2. Логистическая регрессия

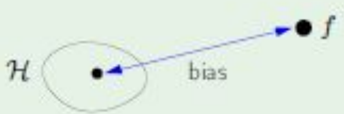
Обзор предыдущей лекции

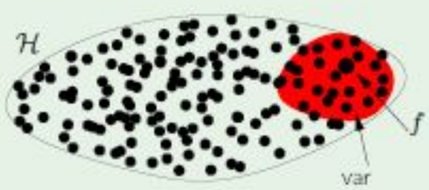
Кривые обучения

- Bias and variance

Expected value of E_{out} w.r.t. $\mathcal{D}$

$$= \text{bias} + \text{var}$$

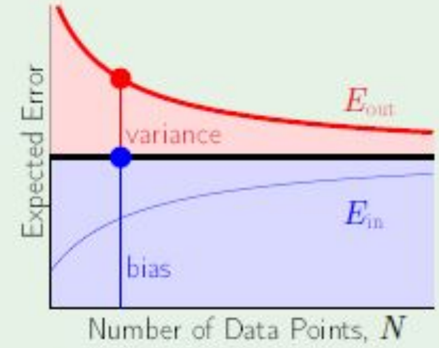




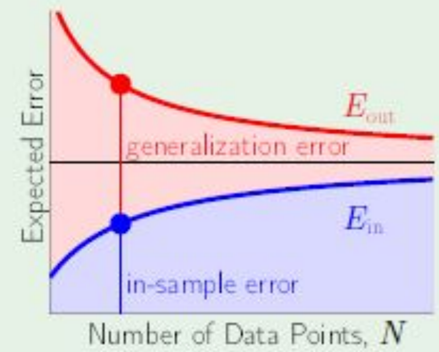
$$g^{(\mathcal{D})}(\mathbf{x}) \rightarrow \bar{g}(\mathbf{x}) \rightarrow f(\mathbf{x})$$

How E_{in} and E_{out} vary with N

B-V:



VC:



- $N \propto$ "VC dimension"

Содержание

- Нелинейное преобразование(обобщение)
- Логистическая регрессия

Нелинейное преобразование

$$\mathbf{x} = (x_0, x_1, \cdots, x_d) \xrightarrow{\Phi} \mathbf{z} = (z_0, z_1, \cdots\cdots\cdots, z_{\tilde{d}})$$

Each $z_i = \phi_i(\mathbf{x})$ $\qquad \mathbf{z} = \Phi(\mathbf{x})$

Example: $\mathbf{z} = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$

Final hypothesis $g(\mathbf{x})$ in $\mathcal{X}$ space:

$\operatorname{sign}\left(\tilde{\mathbf{w}}^\top \Phi(\mathbf{x})\right)$ or $\tilde{\mathbf{w}}^\top \Phi(\mathbf{x})$

Цена, которую мы платим за нелинейное преобразование

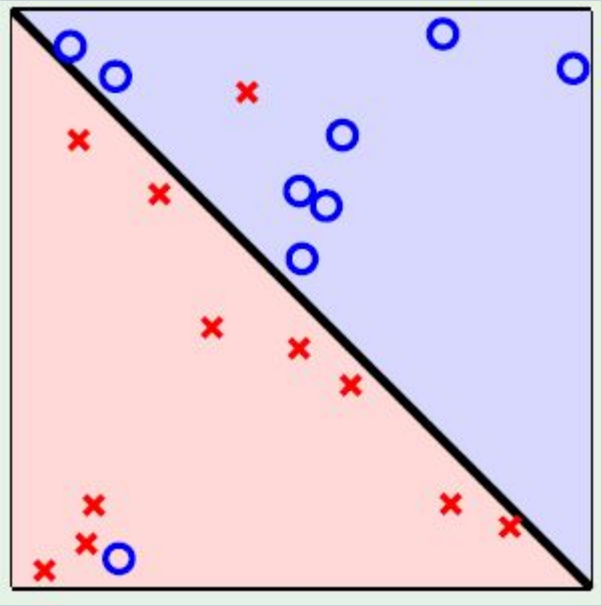
$$\mathbf{x} = (x_0, x_1, \cdots, x_d) \xrightarrow{\Phi} \mathbf{z} = (z_0, z_1, \cdots\cdots\cdots\cdots, z_{\tilde{d}})$$

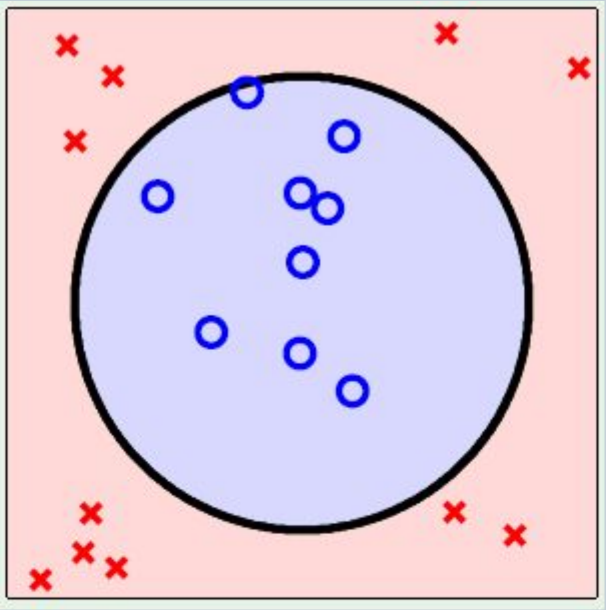
$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$\mathbf{w} \qquad\qquad\qquad\qquad \tilde{\mathbf{w}}$$

$$d_{\mathrm{VC}} = d + 1 \qquad\qquad d_{\mathrm{VC}} \leq \tilde{d} + 1$$

Два линейно неразделимых случая

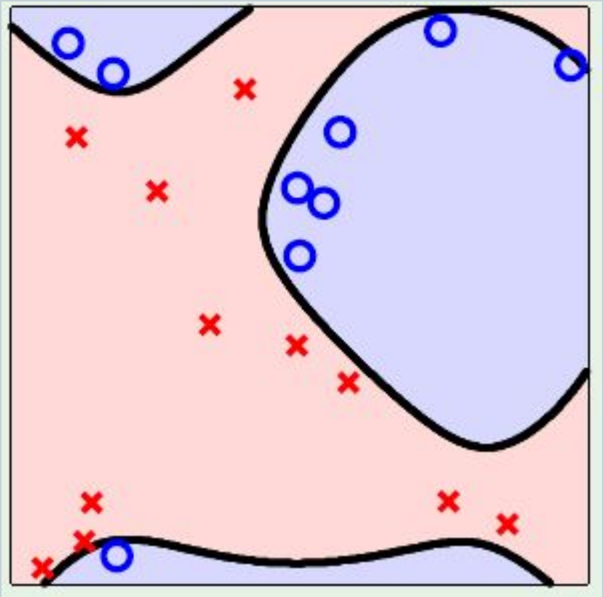




Первый случай

Два способа решения:

- 1) Использовать линейную модель для $E_{in}>0$
- 2) Положить $E_{in}=0$ и использовать нелинейное преобразование в пространстве высокого порядка

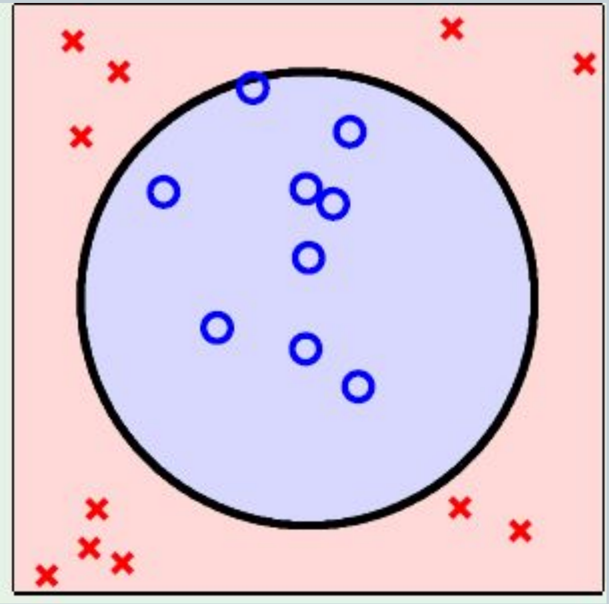


Второй случай

- Перейти к преобразованию

$$\mathbf{z} = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$$

- Или лучше: $\mathbf{z} = (1, x_1^2, x_2^2)$
- Или: $\mathbf{z} = (1, x_1^2 + x_2^2)$
- Или даже: $\mathbf{z} = (x_1^2 + x_2^2 - 0.6)$



Логистическая регрессия

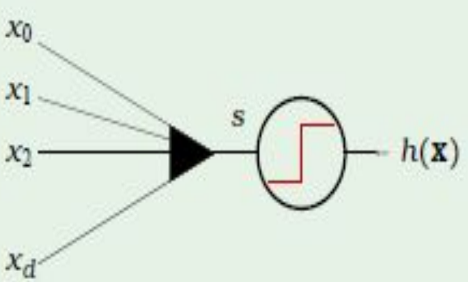
- 1) Модель
- 2) Мера ошибки
- 3) Алгоритм

Линейные модели

$$s = \sum_{i=0}^{d} w_i x_i$$

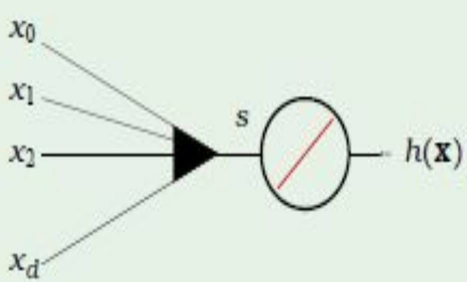
linear classification

$$h(\mathbf{x}) = \mathrm{sign}(s)$$



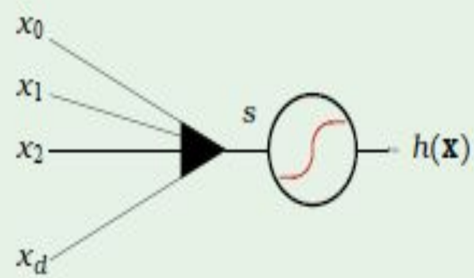
linear regression

$$h(\mathbf{x}) = s$$



logistic regression

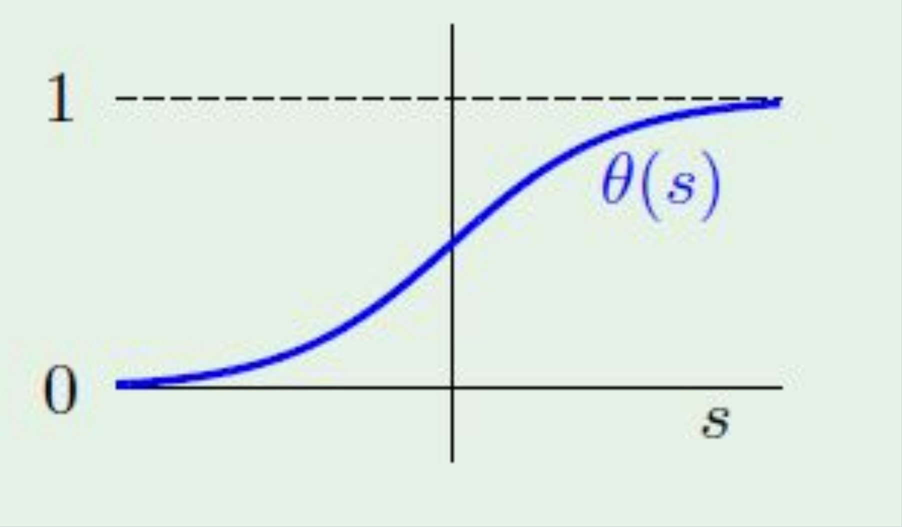
$$h(\mathbf{x}) = \theta(s)$$



Логистическая функция

- «Мягкий порог»

$$\theta(s) = \frac{e^s}{1 + e^s}$$



- Функция θ интерпретируется как вероятность

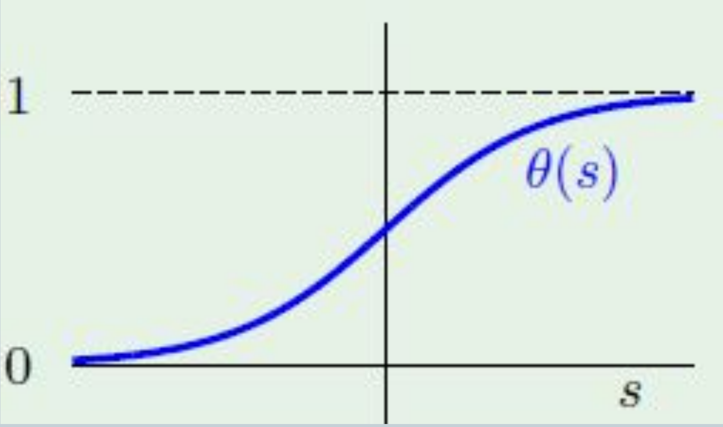
- Данные (x,y), где y бинарная функция, генерируются «шумной» ЦФ

$$P(y \mid \mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{for } y = +1; \\ 1 - f(\mathbf{x}) & \text{for } y = -1. \end{cases}$$

- Функция $f : \mathbb{R}^d \to [0, 1]$ - вероятность
- Изучаем $g(\mathbf{x}) = \theta(\mathbf{w}^\intercal \mathbf{x}) \approx f(\mathbf{x})$

Мера ошибки. Метод максимального правдоподобия

$$P(y \mid \mathbf{x}) = \begin{cases} h(\mathbf{x}) & \text{for } y = +1; \\ 1 - h(\mathbf{x}) & \text{for } y = -1. \end{cases}$$



$$P(y \mid \mathbf{x}) = \theta(y\, \mathbf{w}^{\intercal}\mathbf{x})$$

$$\theta(-s) = 1 - \theta(s)$$

Правдоподобие $\mathcal{D} = (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$ определяется

$$\prod_{n=1}^{N} P(y_n \mid \mathbf{x}_n) = \prod_{n=1}^{N} \theta(y_n \mathbf{w}^{\intercal}\mathbf{x}_n)$$

Максимизируем правдоподобие

- Минимизируем

$$-\frac{1}{N}\ln\left(\prod_{n=1}^{N}\theta(y_n\,\mathbf{w}^{\mathsf{T}}\,\mathbf{x}_n)\right)$$

$$=\frac{1}{N}\sum_{n=1}^{N}\ln\left(\frac{1}{\theta(y_n\,\mathbf{w}^{\mathsf{T}}\,\mathbf{x}_n)}\right)$$

$$E_{\text{in}}(\mathbf{w})=\frac{1}{N}\sum_{n=1}^{N}\underbrace{\ln\left(1+e^{-y_n\mathbf{w}^{\mathsf{T}}\mathbf{x}_n}\right)}_{\mathrm{e}\left(h(\mathbf{x}_n),y_n\right)}$$

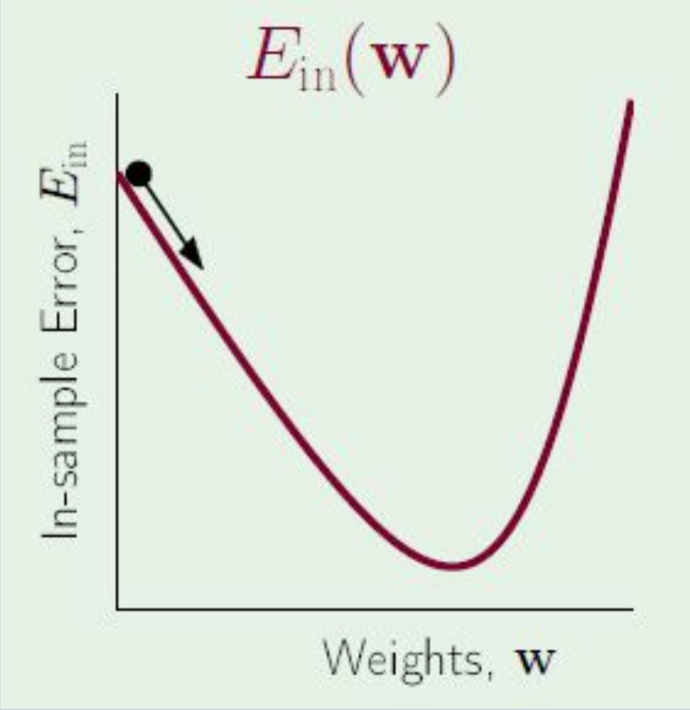
- «кросс-энтропийная» ошибка

Как минимизировать Ein?

$$E_{\text{in}}(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N} \ln\left(1 + e^{-y_n \mathbf{w}^{\mathsf{T}} \mathbf{x}_n}\right)$$

- Метод градиентного спуска

$$\mathbf{w}(1) = \mathbf{w}(0) + \eta\, \hat{\mathbf{v}}$$

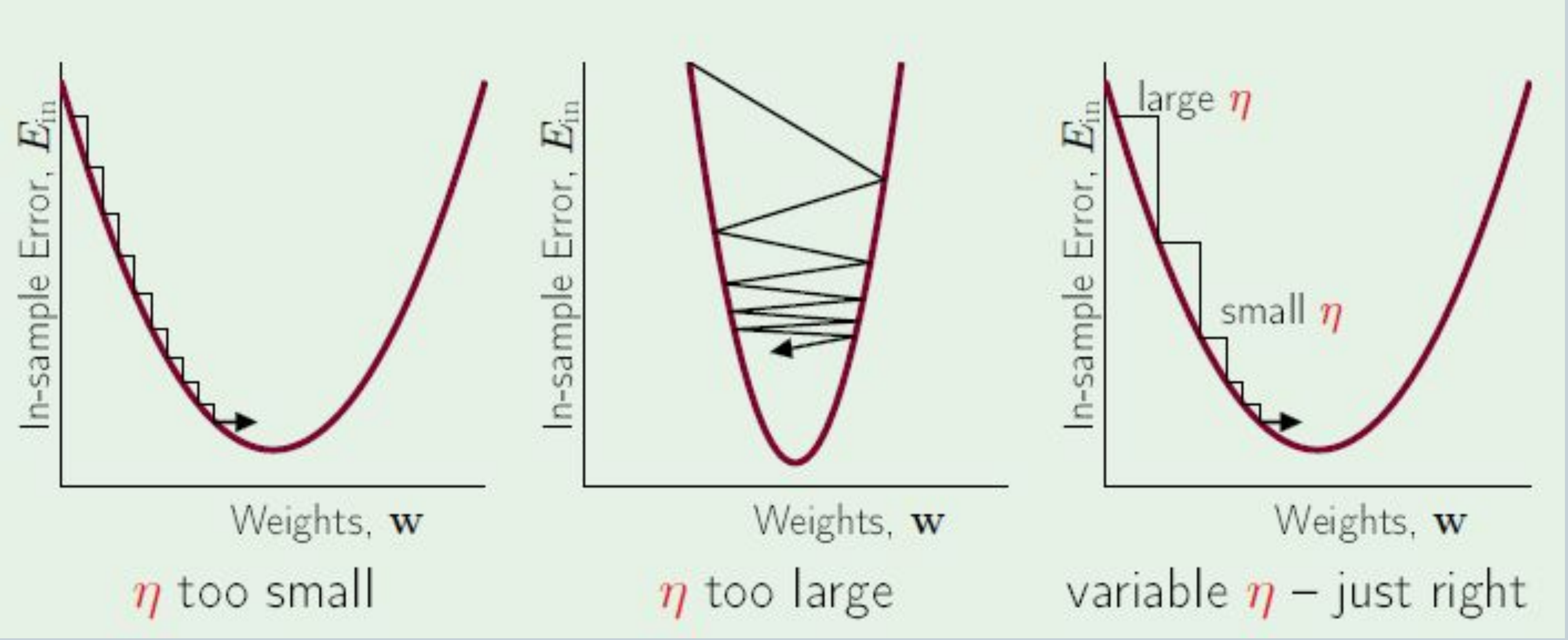


Как выбрать направление $\hat{\mathbf{v}}$

$$
\begin{aligned}
\Delta E_{\text{in}} &= E_{\text{in}}(\mathbf{w}(0) + \eta\hat{\mathbf{v}}) - E_{\text{in}}(\mathbf{w}(0)) \\
&= \eta \nabla E_{\text{in}}(\mathbf{w}(0))^{\mathrm{T}} \hat{\mathbf{v}} + O(\eta^2) \\
&\geq -\eta \|\nabla E_{\text{in}}(\mathbf{w}(0))\|
\end{aligned}
$$

$$
\hat{\mathbf{v}} = -\frac{\nabla E_{\text{in}}(\mathbf{w}(0))}{\|\nabla E_{\text{in}}(\mathbf{w}(0))\|}
$$

Какая длина шага?



Простое применение

- Вместо

$$\Delta \mathbf{w} = \eta \, \hat{\mathbf{v}}$$
$$= -\eta \frac{\nabla E_{\text{in}}(\mathbf{w}(0))}{\|\nabla E_{\text{in}}(\mathbf{w}(0))\|}$$

- Используем

$$\Delta \mathbf{w} = -\eta \, \nabla E_{\text{in}}(\mathbf{w}(0))$$

- С фиксированным η

Алгоритм логистической регрессии

Initialize the weights at $t = 0$ to $\mathbf{w}(0)$

for $t = 0, 1, 2, \ldots$ do

  Compute the gradient

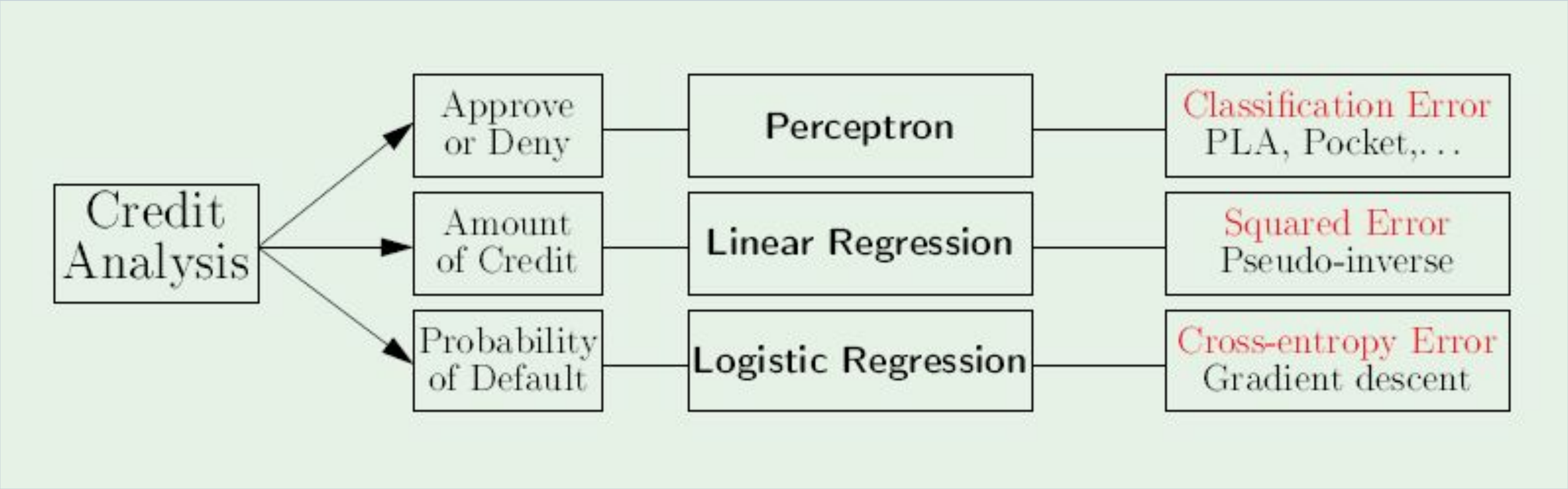
$$\nabla E_{\text{in}} = -\frac{1}{N} \sum_{n=1}^{N} \frac{y_n \mathbf{x}_n}{1 + e^{y_n \mathbf{w}^{\intercal}(t) \mathbf{x}_n}}$$

  Update the weights: $\mathbf{w}(t+1) = \mathbf{w}(t) - \eta \nabla E_{\text{in}}$

  Iterate to the next step until it is time to stop

Return the final weights $\mathbf{w}$

Обзор линейных моделей



Использованные источники

- Yaser Abu Mostafa "Learning from Data: Introductory Machine Learning Course", CalTech University, 2012 . - https://www.youtube.com/playlist?list=PLnIDYuXHkit4LcWjDe0EwlE57WiGlBs08