Семинар 1.

Оценка = 0 × тренир. тесты + 0,2 × Stepic + 0,1 × КР + 0,2 экзамен + 0,5 × ДЗ

Общая задача.

$X = \{x_1, x_2, \ldots, x_n\}$

Обучение с учителем

1) Регрессия $X \to \mathbb{R}$

Номера классов

2) Классификация $X \to 0, 1, \ldots, K$

Без учителя

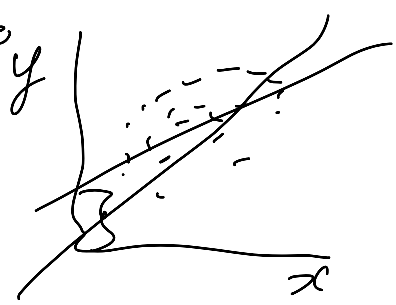
3) Кластеризация

Линейные модели

$$X = \begin{array}{c} \text{признаки} \\ \begin{bmatrix} \underbrace{x_{11} \ldots x_{1k}} \\ \\ x_{n1} \ldots x_{nk} \end{bmatrix} \end{array} \quad \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$y = w_1 x_{11} + \ldots + \underbrace{w_k}\underbrace{x_{1k}}_{+1} + w_0$$

$$y = \langle w, x_1 \rangle$$



~~Признаки~~ Признаки - числовые

Плюсы:

1) Модель простая
2) Она интерпретируема (не всегда)

Скорость в км/ч

Тормозной путь

Вес

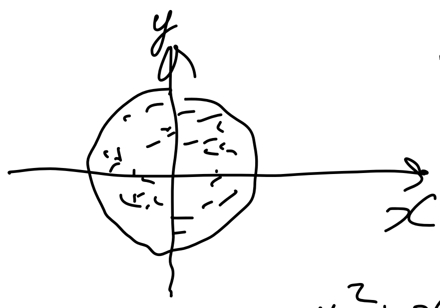
3) Можно усложнять

x_1 - набл.

x^1 - признак

$$y = \underset{1\times1}{w_1} \underset{n\times1}{x^1} + w_2 x^2 + \ldots + w_k x^k + w_0 +$$

$$+ w_{k+1}(x^1)^2 + w_{k+2}(x^2)^3 + \ldots$$



$$Y = w_0 + w_1 x^1$$

$$y^2 + x^2 = C$$

$$y^2 = -w x^2 + c$$

$$y = \cancel{w} -w x^2 + c$$

Минусы:

1) Модель простая
2) Неустойчивость если столбцы линейно зависимы

$$w = \underbrace{\begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_k \end{pmatrix}} \quad X = \underbrace{\begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & & & \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}}_{n \times k}$$

$$\underset{n\times 1}{Y} = \underset{n\times k}{X} \; \underset{k\times 1}{w} \qquad n\times 1 = n \times 1$$

Функционал качества (лосс)

L_2-норма - евклидово расстояние

$$|y - \underbrace{Xw}_{f(X)}|_2$$

$$L(f, X, y) = \|y - Xw\|^2 = \sum_{i=1}^{n} (y_i - \langle x_i, w\rangle)^2$$

$$L(f, X, y) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \langle x_i, w\rangle$$

Mean Squared error

$$(y - Xw)^T (y - Xw) \to \min$$

$$f(x + dx) = f(x) + f'(x)dx + o(dx)$$

$$f(x + dx) = f(x) + L[dx] + o(\|dx\|)$$

$$f(x) : R^n \to R$$

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \qquad df(x) = \langle \nabla_x f(x), dx \rangle$$

$$f(x) : R^{n \times m} \to R$$

$$\nabla_A f(A) = \left(\frac{\partial f}{\partial A_{ij}} \right)_{i,j=1}^{n,m} \qquad tr(\nabla_A^T f(A)\, dA)$$

$$f(x)\quad R^n \to R^m \quad : \quad J = \left(\frac{\partial f_i}{\partial x_j} \right)_{i,j=1}^{n \times m}$$

	скаляр	вектор	матрица
скаляр	$f'(x)dx$	$J\,dx$	—
вектор	$\nabla_x^T f(x) dx$	$J\,dx$	—
матрицы	$tr(\nabla_A f(A)\, dA)$	—	—

$$L = \cancel{\tfrac{1}{n}}(y - Xw)^T(y - Xw) \to \min$$

$$dL(w) = d_w[(\cancel{y} - Xw)^T](y - Xw) + (y - Xw)^T d_w[(\cancel{y} - Xw)] =$$

$$= dw[(-Xw)^T](y - Xw) + (y - Xw)^T dw[-Xw] =$$

$$= -\underbrace{[dw]^T X^T (y - Xw)}_{1\times k \;\; k\times n \;\; n\times 1} - \underbrace{(y - Xw)^T X dw} =$$

$$= -\underbrace{2(y - Xw)^T X}_{\nabla^T} dw$$

$$\nabla_w L = -\frac{2}{n} X^T(y - Xw)$$

$$-\cancel{2}X^T(y - Xw^*) = 0$$

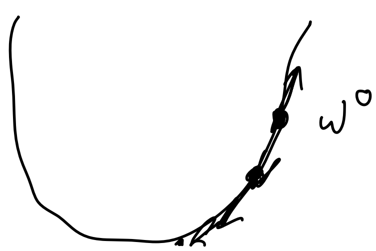
$$X^T y - X^T X w^* = 0 \qquad \to \infty$$

$$X^T y = X^T X w^* \qquad w^* = \underbrace{(X^T X)^{-1} X^T y}$$

$$E(Y|X) = Xw^* \qquad \frac{1}{|X^T X|} \to \infty$$

числа обусловленности

$L(w)$



$$w^0 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \qquad w \sim N(0, I)$$

$$j > i \qquad = \begin{pmatrix} \frac{\partial L}{\partial w_1} \\ \vdots \\ \frac{\partial L}{\partial w_k} \end{pmatrix}$$

$$\underset{k \times 1}{w_j} = \underset{k \times 1}{w_i} - \alpha \nabla_w L(w_i)$$

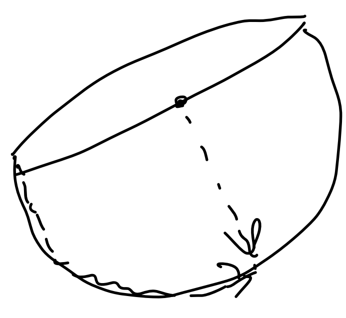
$$w_1 = w_0 + \ldots$$

$$\nabla_w L = -\frac{2}{n} X^T (y - Xw)$$

$$w_j = w_i + \alpha \frac{2}{n} X^T (y - X w_i)$$

1) Выпуклая, выбор старта влияет, но не сильно

2)



3) α сильно влияет.