

# Разбор летучки

---

# Лекция 10

## Линейная регрессия

---

Екатерина Тузова

$X$  — объекты в  $\mathbb{R}^n$ ;  $Y$  — ответы в  $\mathbb{R}$

$X^l = (x_i, y_i)_{i=1}^l$  — обучающая выборка

$y_i = y(x_i)$ ,  $y : X \rightarrow Y$  — неизвестная зависимость

$X$  — объекты в  $\mathbb{R}^n$ ;  $Y$  — ответы в  $\mathbb{R}$

$X^l = (x_i, y_i)_{i=1}^l$  — обучающая выборка

$y_i = y(x_i)$ ,  $y : X \rightarrow Y$  — неизвестная зависимость

$a(x) = f(x, w)$  — модель зависимости,

$w \in \mathbb{R}^p$  — вектор параметров модели.

# Многомерная линейная регрессия

$x^1, \dots, x^n$  – числовые признаки

Модель многомерной линейной регрессии:

$$f(x, w) = \sum_{j=1}^n w_j x^j, \quad w \in \mathbb{R}$$

# Многомерная линейная регрессия

$x^1, \dots, x^n$  – числовые признаки

Модель многомерной линейной регрессии:

$$f(x, w) = \sum_{j=1}^n w_j x^j, \quad w \in \mathbb{R}$$

Функционал квадрата ошибки:

$$Q(w, X^l) = \sum_{i=1}^l (f(x_i, w) - y_i)^2 \rightarrow \min_w$$

# Матричное представление

$$X_{l \times n} = \begin{pmatrix} x_1^1 & \dots & x_1^n \\ \dots & \dots & \dots \\ x_l^1 & \dots & x_l^n \end{pmatrix} \quad y_{l \times 1} = \begin{pmatrix} y_1 \\ \dots \\ y_l \end{pmatrix} \quad w_{n \times 1} = \begin{pmatrix} w_1 \\ \dots \\ w_n \end{pmatrix}$$

# Матричное представление

$$X_{l \times n} = \begin{pmatrix} x_1^1 & \dots & x_1^n \\ \dots & \dots & \dots \\ x_l^1 & \dots & x_l^n \end{pmatrix} \quad y_{l \times 1} = \begin{pmatrix} y_1 \\ \dots \\ y_l \end{pmatrix} \quad w_{n \times 1} = \begin{pmatrix} w_1 \\ \dots \\ w_n \end{pmatrix}$$

Функционал квадрата ошибки:

$$Q(w, X^l) = \sum_{i=1}^l (f(x_i, w) - y_i)^2 = \|Xw - y\|^2 \rightarrow \min_w$$



Необходимое условие минимума:

$$\frac{\partial Q(w)}{\partial w} = 2X^T(Xw - y) = 0$$

# Нормальная система уравнений

Необходимое условие минимума:

$$\frac{\partial Q(w)}{\partial w} = 2X^T(Xw - y) = 0$$

Откуда следует нормальная система задачи МНК:

$$X^T X w = X^T y$$

$X^T X$  – ковариационная матрица признаков  $x^1, \dots, x^n$

# Нормальная система уравнений

Нормальная система задачи МНК:

$$X^T X w = X^T y$$

# Нормальная система уравнений

Нормальная система задачи МНК:

$$X^T X w = X^T y$$

Решение системы:

$$w^* = (X^T X)^{-1} X^T y = X^+ y$$

$X^+$  – псевдообратная матрица

# Нормальная система уравнений

Нормальная система задачи МНК:

$$X^T X w = X^T y$$

Решение системы:

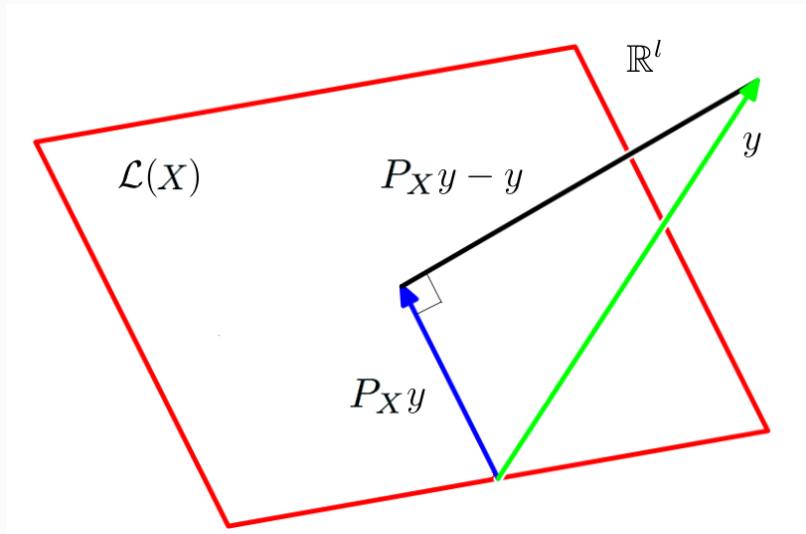
$$w^* = (X^T X)^{-1} X^T y = X^+ y$$

$X^+$  – псевдообратная матрица

Значение функционала:  $Q(w^*) = \|P_X y - y\|^2$

где  $P_X = X X^+ = X (X^T X)^{-1} X^T$  – проекционная матрица

# Геометрический смысл



# Сингулярное разложение (экономное)

Произвольная  $l \times n$ -матрица представима в виде сингулярного разложения:

$$X = VDU^T$$

# Сингулярное разложение (экономное)

Произвольная  $l \times n$ -матрица представима в виде сингулярного разложения:

$$X = VDU^T$$

Основные свойства сингулярного разложения:

- $V_{l \times n} = (v_1, \dots, v_n)$  ортогональна,  $V^T V = I_n$ ,  
столбцы  $v_j$  - собственные векторы матрицы  $XX^T$
- $U_{n \times n} = (u_1, \dots, u_n)$  ортогональна,  $U^T U = I_n$ ,  
столбцы  $u_j$  - собственные векторы матрицы  $X^T X$
- $D$  диагональна,  $D_{n \times n} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ ,  
 $\lambda_j > 0$  - собственные значения матриц  $X^T X$  и  $XX^T$



## Решение МНК через сингулярное разложение

Псевдообратная  $X^+$ , вектор МНК-решения  $w^*$ , МНК-аппроксимация целевого вектора  $Xw^*$

Псевдообратная  $X^+$ , вектор МНК-решения  $w^*$ , МНК-аппроксимация целевого вектора  $Xw^*$

$$X^+ = (UDV^TVDU^T)^{-1}UDV^T = UD^{-1}V^T = \sum_{j=1}^n \frac{1}{\sqrt{\lambda_j}} u_j v_j^T$$

## Решение МНК через сингулярное разложение

Псевдообратная  $X^+$ , вектор МНК-решения  $w^*$ , МНК-аппроксимация целевого вектора  $Xw^*$

$$X^+ = (UDV^TVDU^T)^{-1}UDV^T = UD^{-1}V^T = \sum_{j=1}^n \frac{1}{\sqrt{\lambda_j}} u_j v_j^T$$
$$w^* = X^+y = UD^{-1}V^T y = \sum_{j=1}^n \frac{1}{\sqrt{\lambda_j}} u_j (v_j^T y)$$

## Решение МНК через сингулярное разложение

Псевдообратная  $X^+$ , вектор МНК-решения  $w^*$ , МНК-аппроксимация целевого вектора  $Xw^*$

$$X^+ = (UDV^TVDU^T)^{-1}UDV^T = UD^{-1}V^T = \sum_{j=1}^n \frac{1}{\sqrt{\lambda_j}} u_j v_j^T$$

$$w^* = X^+y = UD^{-1}V^T y = \sum_{j=1}^n \frac{1}{\sqrt{\lambda_j}} u_j (v_j^T y)$$

$$Xw^* = P_X y = (VDU^T)UD^{-1}V^T y = VV^T y = \sum_{j=1}^n v_j (v_j^T y)$$

## Решение МНК через сингулярное разложение

Псевдообратная  $X^+$ , вектор МНК-решения  $w^*$ , МНК-аппроксимация целевого вектора  $Xw^*$

$$X^+ = (UDV^TVDU^T)^{-1}UDV^T = UD^{-1}V^T = \sum_{j=1}^n \frac{1}{\sqrt{\lambda_j}} u_j v_j^T$$

$$w^* = X^+y = UD^{-1}V^T y = \sum_{j=1}^n \frac{1}{\sqrt{\lambda_j}} u_j (v_j^T y)$$

$$Xw^* = P_X y = (VDU^T)UD^{-1}V^T y = VV^T y = \sum_{j=1}^n v_j (v_j^T y)$$

$$\|w^*\|^2 = \|UD^{-1}V^T y\|^2 = \sum_{j=1}^n \frac{1}{\lambda_j} (v_j^T y)^2$$

Если матрица  $X^T X$  плохо обусловлена, то:

- решение становится неустойчивым и неинтерпретируемым,  $\|w^*\|$  велико
- $E_{in} = \|Xw^* - y\|$  - мало
- $E_{out} = \|X'w^* - y'\|$  - велико

Если матрица  $X^T X$  плохо обусловлена, то:

- решение становится неустойчивым и неинтерпретируемым,  $\|w^*\|$  велико
- $E_{in} = \|Xw^* - y\|$  – мало
- $E_{out} = \|X'w^* - y'\|$  – велико

**Вопрос:** Как бороться с этой проблемой?

Стратегии устранения мультиколлинеарности и переобучения:

- Регуляризация:  $\|w\| \rightarrow \min$
- Отбор признаков:  $x^1, \dots, x^n \rightarrow x^{j_1}, \dots, x^{j_m}, \quad m \ll n$
- Преобразование признаков:  $x^1, \dots, x^n \rightarrow g^1, \dots, g^m, \quad m \ll n$



Штраф за увеличение нормы вектора весов  $\|w\|$ :

$$Q_\tau(w) = \|Xw - y\|^2 + \tau\|w\|^2$$

где  $\tau$  – неотрицательный параметр регуляризации.

Штраф за увеличение нормы вектора весов  $\|w\|$ :

$$Q_\tau(w) = \|Xw - y\|^2 + \tau\|w\|^2$$

где  $\tau$  – неотрицательный параметр регуляризации.

Модифицированное МНК-решение ( $\tau I_n$  – «гребень»)

$$w_\tau^* = (X^T X + \tau I_n)^{-1} X^T y$$

Штраф за увеличение нормы вектора весов  $\|w\|$ :

$$Q_\tau(w) = \|Xw - y\|^2 + \tau\|w\|^2$$

где  $\tau$  – неотрицательный параметр регуляризации.

Модифицированное МНК-решение ( $\tau I_n$  – «гребень»)

$$w_\tau^* = (X^T X + \tau I_n)^{-1} X^T y$$

**Вопрос:** Можно ли подбирать  $\tau$  не вычисляя каждый раз обратную матрицу?

Модифицированное МНК-решение ( $\tau I_n$  — «гребень»)

$$w_{\tau}^* = (X^T X + \tau I_n)^{-1} X^T y$$

Модифицированное МНК-решение ( $\tau I_n$  — «гребень»)

$$w_{\tau}^* = (X^T X + \tau I_n)^{-1} X^T y$$

Преимущество сингулярного разложения:

Можно подбирать параметр  $\tau$ , вычислив сингулярное разложение только один раз.

$$w_{\tau}^* = U(D^2 + \tau I_n)^{-1} D V^T y = \sum_{j=1}^n \frac{\sqrt{\lambda_j}}{\lambda_j + \tau} u_j (v_j^T y)$$

$$w_{\tau}^* = U(D^2 + \tau I_n)^{-1} D V^T y = \sum_{j=1}^n \frac{\sqrt{\lambda_j}}{\lambda_j + \tau} u_j (v_j^T y)$$

$$X w_{\tau}^* = V D U^T w_{\tau}^* = V \operatorname{diag} \left( \frac{\lambda_j}{\lambda_j + \tau} \right) V^T y = \sum_{j=1}^n \frac{\lambda_j}{\lambda_j + \tau} v_j (v_j^T y)$$

$$w_\tau^* = U(D^2 + \tau I_n)^{-1} D V^T y = \sum_{j=1}^n \frac{\sqrt{\lambda_j}}{\lambda_j + \tau} u_j (v_j^T y)$$

$$X w_\tau^* = V D U^T w_\tau^* = V \operatorname{diag} \left( \frac{\lambda_j}{\lambda_j + \tau} \right) V^T y = \sum_{j=1}^n \frac{\lambda_j}{\lambda_j + \tau} v_j (v_j^T y)$$

$$\|w_\tau^*\|^2 = \|U(D^2 + \tau I_n)^{-1} D V^T y\|^2 = \sum_{j=1}^n \frac{\lambda_j}{(\lambda_j + \tau)^2} (v_j^T y)^2$$



$$w_\tau^* = U(D^2 + \tau I_n)^{-1} D V^T y = \sum_{j=1}^n \frac{\sqrt{\lambda_j}}{\lambda_j + \tau} u_j (v_j^T y)$$

$$X w_\tau^* = V D U^T w_\tau^* = V \operatorname{diag} \left( \frac{\lambda_j}{\lambda_j + \tau} \right) V^T y = \sum_{j=1}^n \frac{\lambda_j}{\lambda_j + \tau} v_j (v_j^T y)$$

$$\|w_\tau^*\|^2 = \|U(D^2 + \tau I_n)^{-1} D V^T y\|^2 = \sum_{j=1}^n \frac{\lambda_j}{(\lambda_j + \tau)^2} (v_j^T y)^2$$

$X w_\tau^* \neq X w^*$  – зато решение становится более устойчивым

Контрольная выборка:  $X^k = (x'_i, y'_i)_{i=1}^k$

$$Q(w_\tau^*, X^k) = \|X'w_\tau^* - y'\|^2 = \|X'U \operatorname{diag}\left(\frac{\lambda_j}{\lambda_j + \tau}\right) V^T y - y'\|^2$$

## Выбор параметра регуляризации $\tau$

Контрольная выборка:  $X^k = (x'_i, y'_i)_{i=1}^k$

$$Q(w_\tau^*, X^k) = \|X'w_\tau^* - y'\|^2 = \|X'U \operatorname{diag}\left(\frac{\lambda_j}{\lambda_j + \tau}\right) V^T y - y'\|^2$$

Зависимость  $Q(\tau)$  обычно имеет характерный минимум.

$$\begin{cases} Q(w, X^l) = \|Xw - y\|^2 \rightarrow \min_w \\ \sum_{j=1}^n |w_j| \leq \kappa \end{cases}$$

$$\begin{cases} Q(w, X^l) = \|Xw - y\|^2 \rightarrow \min_w \\ \sum_{j=1}^n |w_j| \leq \kappa \end{cases}$$

После замены переменных:

$$\begin{cases} w_j = w_j^+ - w_j^- \\ |w_j| = w_j^+ + w_j^-, \quad w_j^+, w_j^- \geq 0 \end{cases}$$

ограничения принимают канонический вид:

$$\sum_{j=1}^n w_j^+ + w_j^- \leq \kappa$$

$$\begin{cases} Q(w, X^l) = \|Xw - y\|^2 \rightarrow \min_w \\ \sum_{j=1}^n |w_j| \leq \kappa \end{cases}$$

После замены переменных:

$$\begin{cases} w_j = w_j^+ - w_j^- \\ |w_j| = w_j^+ + w_j^-, \quad w_j^+, w_j^- \geq 0 \end{cases}$$

ограничения принимают канонический вид:

$$\sum_{j=1}^n w_j^+ + w_j^- \leq \kappa$$

Чем меньше  $\kappa$ , тем больше  $j$  таких, что  $w_j^+ = w_j^- = 0$

# Нелинейная регрессия

---

Вопрос: Что изменится, если модель регрессии не линейна?

$$f(x, w), \quad w \in \mathbb{R}^p$$



Начальное приближение  $w^{(0)} = (w_1^{(0)}, \dots, w_p^{(0)})$

Итерационный процесс:  $w^{(t+1)} = w^{(t)} - h_t(Q''(w^{(t)}))^{-1}Q'(w^{(t)})$

$Q'(w^{(t)})$  – градиент функционала  $Q$  в точке  $w^{(t)}$

$Q''(w^{(t)})$  – гессиан функционала  $Q$  в точке  $w^{(t)}$

$h_t$  – величина шага

$$Q(w, X^l) = \sum_{i=1}^l (f(x_i, w) - y_i)^2 \rightarrow \min_w$$

$$Q(w, X^l) = \sum_{i=1}^l (f(x_i, w) - y_i)^2 \rightarrow \min_w$$

$$\frac{\partial Q(w)}{\partial w_j} = 2 \sum_{i=1}^l (f(x_i, w) - y_i) \frac{\partial (f(x_i, w))}{\partial w_j}$$

$$Q(w, X^l) = \sum_{i=1}^l (f(x_i, w) - y_i)^2 \rightarrow \min_w$$

$$\frac{\partial Q(w)}{\partial w_j} = 2 \sum_{i=1}^l (f(x_i, w) - y_i) \frac{\partial(f(x_i, w))}{\partial w_j}$$

$$\frac{\partial^2 Q(w)}{\partial w_j \partial w_k} = 2 \sum_{i=1}^l \frac{\partial(f(x_i, w))}{\partial w_j} \frac{\partial(f(x_i, w))}{\partial w_k} - 2 \sum_{i=1}^l (f(x_i, w) - y_i) \frac{\partial^2(f(x_i, w))}{\partial w_j \partial w_k}$$

$$Q(w, X^l) = \sum_{i=1}^l (f(x_i, w) - y_i)^2 \rightarrow \min_w$$

$$\frac{\partial Q(w)}{\partial w_j} = 2 \sum_{i=1}^l (f(x_i, w) - y_i) \frac{\partial (f(x_i, w))}{\partial w_j}$$

$$\frac{\partial^2 Q(w)}{\partial w_j \partial w_k} = 2 \sum_{i=1}^l \frac{\partial (f(x_i, w))}{\partial w_j} \frac{\partial (f(x_i, w))}{\partial w_k} - 2 \sum_{i=1}^l (f(x_i, w) - y_i) \frac{\partial^2 (f(x_i, w))}{\partial w_j \partial w_k}$$

**Вопрос:** Какая часть самая тяжелая?

$$\frac{\partial Q(w)}{\partial w_j} = 2 \sum_{i=1}^l (f(x_i, w) - y_i) \frac{\partial (f(x_i, w))}{\partial w_j}$$

$$\frac{\partial Q(w)}{\partial w_j} = 2 \sum_{i=1}^l (f(x_i, w) - y_i) \frac{\partial(f(x_i, w))}{\partial w_j}$$

$$\frac{\partial^2 Q(w)}{\partial w_j \partial w_k} = 2 \sum_{i=1}^l \frac{\partial(f(x_i, w))}{\partial w_j} \frac{\partial(f(x_i, w))}{\partial w_k} - 2 \sum_{i=1}^l (f(x_i, w) - y_i) \frac{\partial^2(f(x_i, w))}{\partial w_j \partial w_k}$$

$$\frac{\partial Q(w)}{\partial w_j} = 2 \sum_{i=1}^l (f(x_i, w) - y_i) \frac{\partial(f(x_i, w))}{\partial w_j}$$

$$\frac{\partial^2 Q(w)}{\partial w_j \partial w_k} = 2 \sum_{i=1}^l \frac{\partial(f(x_i, w))}{\partial w_j} \frac{\partial(f(x_i, w))}{\partial w_k} - 2 \sum_{i=1}^l (f(x_i, w) - y_i) \frac{\partial^2(f(x_i, w))}{\partial w_j \partial w_k}$$

Линеаризация  $f(x_i, w)$  в окрестности текущего  $w^{(t)}$ :

$$f(x_i, w) = f(x_i, w^{(t)}) + \sum_{j=1}^p \frac{\partial(f(x_i, w))}{\partial w_j} (w_j - w_j^{(t)}) + o(w_j - w_j^{(t)})$$

⇒ второе слагаемое в гессиане обнулилось



# Матричные обозначения

$$X_t = \left( \frac{\partial(f(x_i, w^{(t)}))}{\partial w_j^{(t)}} \right)_{l \times p} \quad \text{– матрица первых производных}$$
$$f_t = (f(x_i, w^{(t)}))_{l \times 1} \quad \text{– вектор значений } f$$

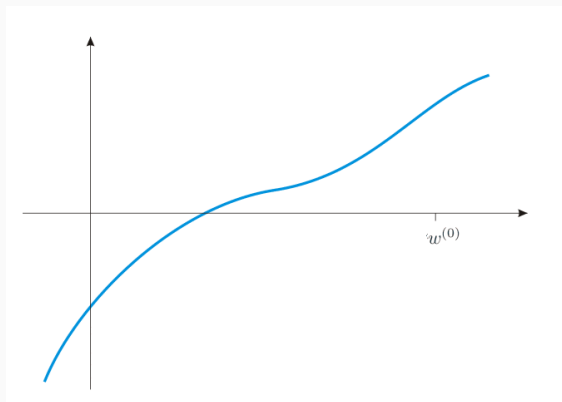
Формула  $t$ -й итерации метода Ньютона–Гаусса:

$$w^{(t+1)} = w^{(t)} - h_t \underbrace{(X_t^T X_t)^{-1} X_t^T (f_t - y)}_{\beta}$$

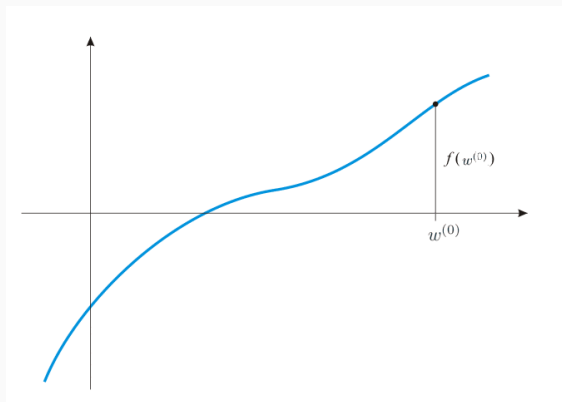
где  $\beta$  – решение многомерной линейной регрессии

$$\|X_t \beta - (f_t - y)\|^2 \rightarrow \min_{\beta}$$

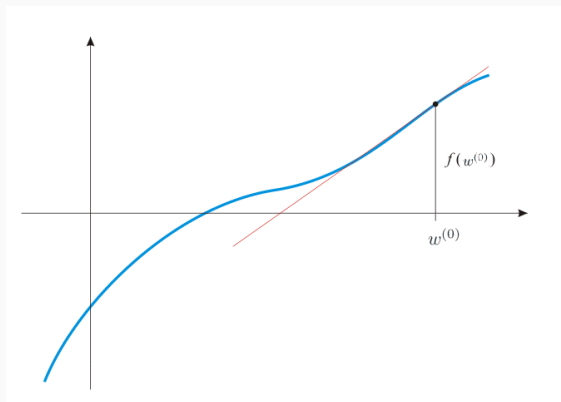
# Метод Ньютона-Рафсена



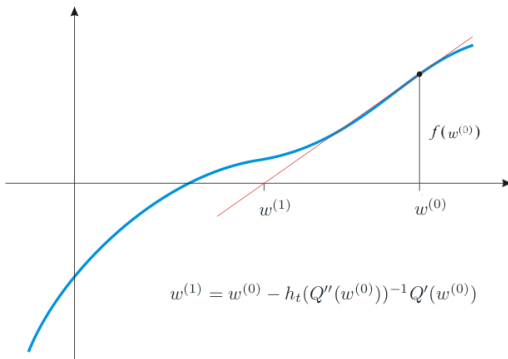
# Метод Ньютона-Рафсена



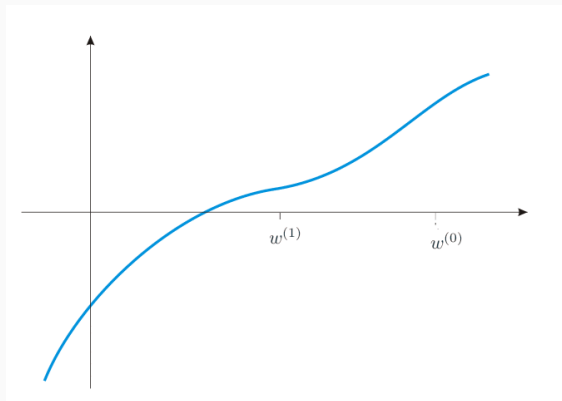
# Метод Ньютона-Рафсена



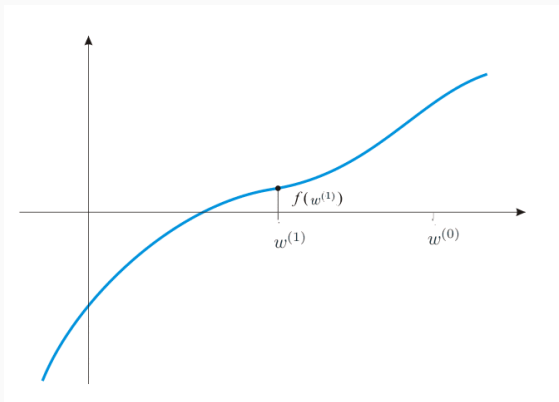
# Метод Ньютона-Рафсена



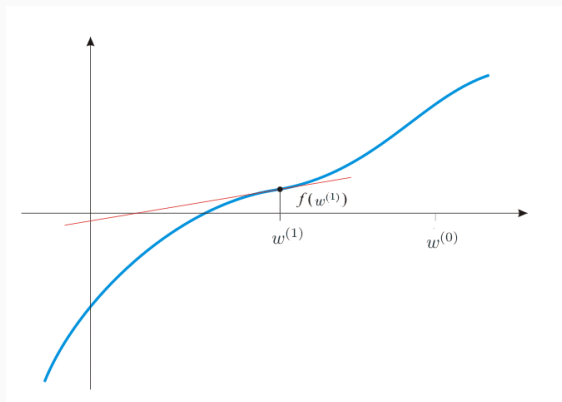
# Метод Ньютона-Рафсена



# Метод Ньютона-Рафсена

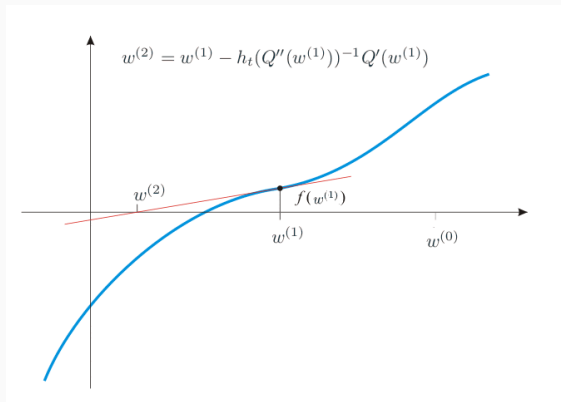


# Метод Ньютона-Рафсена

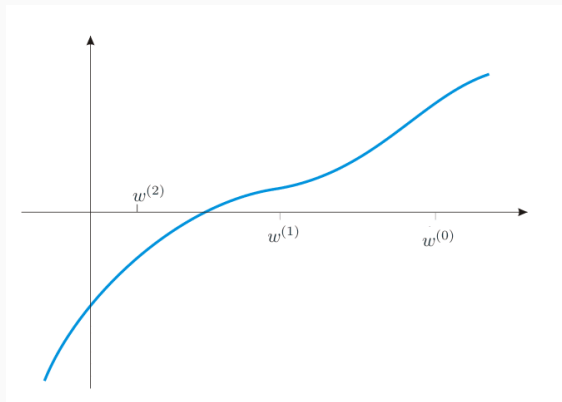




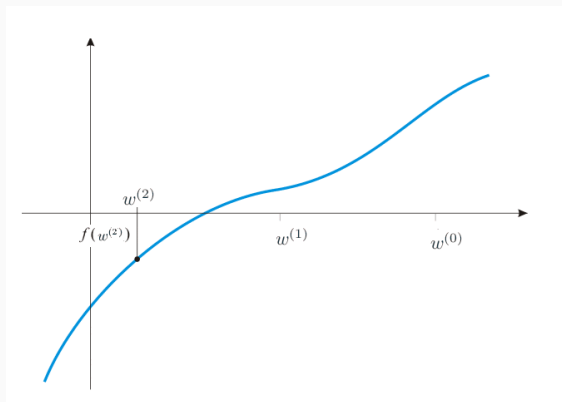
# Метод Ньютона-Рафсена



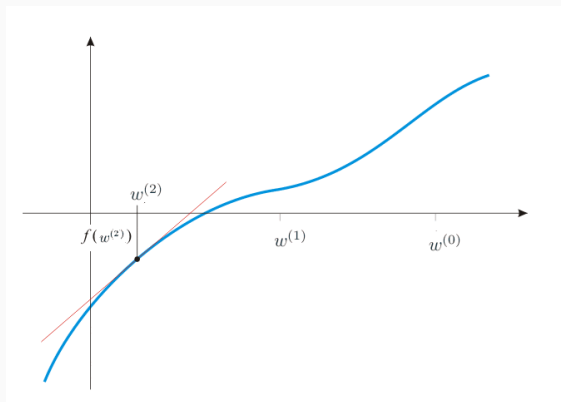
# Метод Ньютона-Рафсена



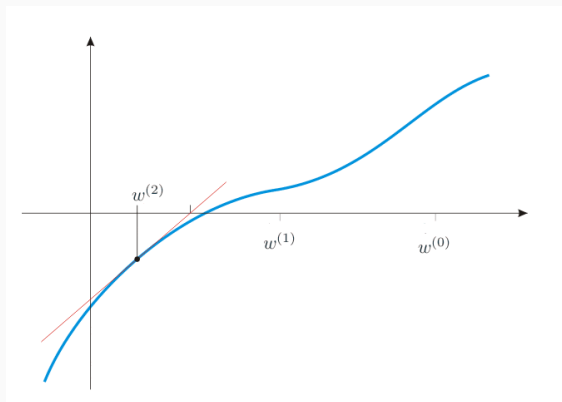
# Метод Ньютона-Рафсена



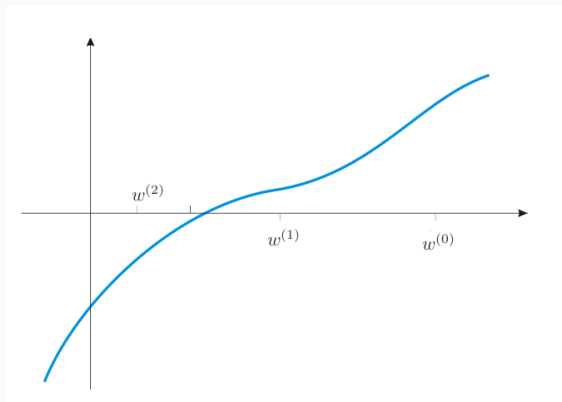
# Метод Ньютона-Рафсена



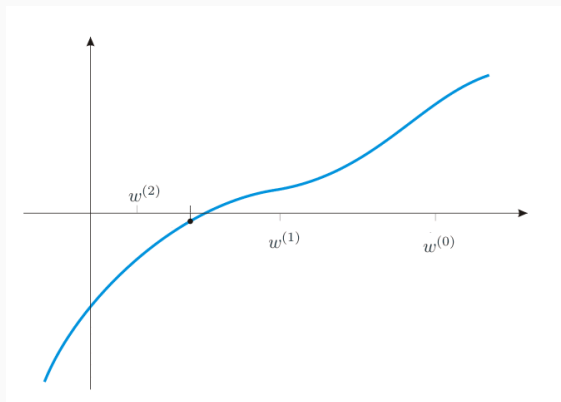
# Метод Ньютона-Рафсена



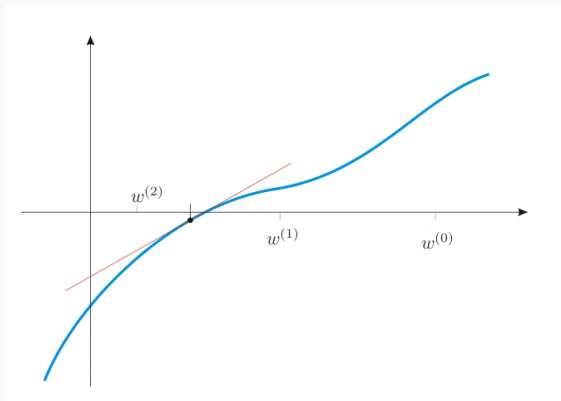
# Метод Ньютона-Рафсена



# Метод Ньютона-Рафсена

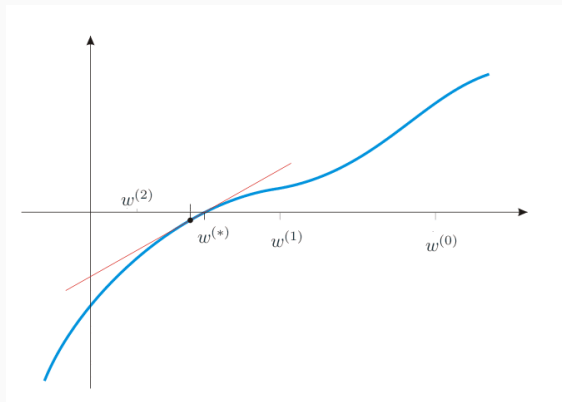


# Метод Ньютона-Рафсена





# Метод Ньютона-Рафсена



Вопросы?

# Метод главных компонент

---

$x^1, \dots, x^n$  – исходные числовые признаки

$g^1(x), \dots, g^m(x)$  – новые числовые признаки,  $m \times n$

**Вопрос:** Как сформулировать требование к новым признакам?

# Метод главных компонент

$x^1, \dots, x^n$  – исходные числовые признаки

$g_1(x), \dots, g_m(x)$  – новые числовые признаки,  $m \times n$

# Метод главных компонент

$x^1, \dots, x^n$  – исходные числовые признаки

$g_1(x), \dots, g_m(x)$  – новые числовые признаки,  $m \times n$

Требование: старые признаки должны линейно восстанавливаться по новым:

$$\hat{x}^j = \sum_{s=1}^m g_s(x) u_{js}, \quad j = 1, \dots, n, \quad \forall x \in X$$

как можно точнее на обучающей выборке  $x_1, \dots, x_l$ :

$$\sum_{i=1}^l \sum_{j=1}^n (\hat{x}_i^j - x_i)^2 \rightarrow \min_{g_s(x_i), u_{js}}$$

## Матричные обозначения

$$X_{l \times n} = \begin{pmatrix} x_1^1 & \dots & x_1^n \\ \dots & \dots & \dots \\ x_l^1 & \dots & x_l^n \end{pmatrix} \quad G_{l \times m} = \begin{pmatrix} g_1(x_1) & \dots & g_m(x_1) \\ \dots & \dots & \dots \\ g_1(x_l) & \dots & g_m(x_l) \end{pmatrix}$$

$$U_{n \times m} = \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \dots & \dots & \dots \\ u_{n1} & \dots & u_{nm} \end{pmatrix}$$

$U$  – линейное преобразование новых признаков в старые

$$\hat{X} = GU^T \approx X$$

## Матричные обозначения

$$X_{l \times n} = \begin{pmatrix} x_1^1 & \dots & x_1^n \\ \dots & \dots & \dots \\ x_l^1 & \dots & x_l^n \end{pmatrix} \quad G_{l \times m} = \begin{pmatrix} g_1(x_1) & \dots & g_m(x_1) \\ \dots & \dots & \dots \\ g_1(x_l) & \dots & g_m(x_l) \end{pmatrix}$$
$$U_{n \times m} = \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \dots & \dots & \dots \\ u_{n1} & \dots & u_{nm} \end{pmatrix}$$

$U$  – линейное преобразование новых признаков в старые

$$\hat{X} = GU^T \approx X$$

**Найти:** новые признаки  $G$  и преобразование  $U$ :

$$\sum_{i=1}^l \sum_{j=1}^n (\hat{x}_i^j - x_i^j)^2 = \|GU^T - X\|^2 \rightarrow \min_{G,U}$$



# Основная теорема

Если  $m < \text{rank } X$ , то минимум  $\|GU^T - X\|^2$  достигается, когда столбцы  $U$  - это с.в. матрицы  $X^T X$ , соответствующие  $m$  максимальным с.з.  $\lambda_1, \dots, \lambda_m$ , а матрица  $G = XU$ .

# Основная теорема

Если  $m < \text{rank } X$ , то минимум  $\|GU^T - X\|^2$  достигается, когда столбцы  $U$  - это с.в. матрицы  $X^T X$ , соответствующие  $m$  максимальным с.з.  $\lambda_1, \dots, \lambda_m$ , а матрица  $G = XU$ .

При этом:

- матрица  $U$  ортонормирована:  $U^T U = I_m$
- матрица  $G$  ортогональна:  $G^T G = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$
- $U\Lambda = X^T XU$ ,  $G\Lambda = FF^T G$
- $\|GU^T - X\|^2 = \|X\|^2 - \text{tr}\Lambda = \sum_{j=m+1}^n \lambda_j$

## Связь с сингулярным разложением

Если взять  $m = n$ , то:

$$\cdot \|GU^T - X\|^2 = 0$$

## Связь с сингулярным разложением

Если взять  $m = n$ , то:

- $\|GU^T - X\|^2 = 0$
- представление  $\hat{X} = GU^T = X$  точное и совпадает с сингулярным разложением при  $G = V\sqrt{\Lambda}$

$$X = GU^T = V\sqrt{\Lambda}U^T, \quad U^TU = I_m, \quad V^TV = I_m$$

## Связь с сингулярным разложением

Если взять  $m = n$ , то:

- $\|GU^T - X\|^2 = 0$
- представление  $\hat{X} = GU^T = X$  точное и совпадает с сингулярным разложением при  $G = V\sqrt{\Lambda}$

$$X = GU^T = V\sqrt{\Lambda}U^T, \quad U^TU = I_m, \quad V^TV = I_m$$

- линейное преобразование  $U$  работает в обе стороны:

$$X = GU^T, \quad G = XU$$

Преобразование  $U$  называется декоррелирующим

## Эффективная размерность выборки

Упорядочим с.з.  $X^T X$  по убыванию:  $\lambda_1 > \dots > \lambda_n > 0$

Эффективная размерность выборки – это наименьшее целое  $m$ , при котором

$$E_m = \frac{\|GU^T - X\|^2}{\|X\|^2} = \frac{\lambda_{m+1} + \dots + \lambda_n}{\lambda_1 + \dots + \lambda_n} \leq \varepsilon$$

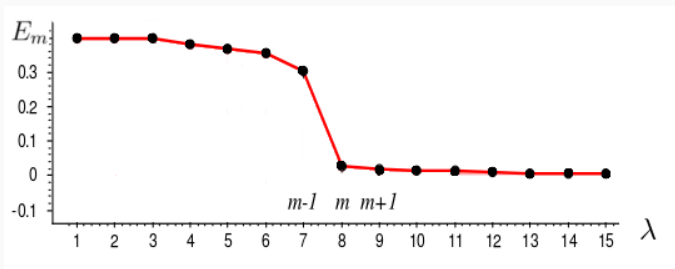
# Эффективная размерность выборки

Упорядочим с.з.  $X^T X$  по убыванию:  $\lambda_1 > \dots > \lambda_n > 0$

Эффективная размерность выборки – это наименьшее целое  $m$ , при котором

$$E_m = \frac{\|GU^T - X\|^2}{\|X\|^2} = \frac{\lambda_{m+1} + \dots + \lambda_n}{\lambda_1 + \dots + \lambda_n} \leq \varepsilon$$

Критерий «крутого склона»: находим  $m : E_{m-1} \gg E_m$ :



## Решение задачи НК в новых признаках

Заменяем  $F$  на её приближение  $GU^T$ :

$$\|GU^T w - y\|^2 = \|G\hat{w} - y\|^2 \rightarrow \min_{\hat{w}}$$

Связь нового и старого вектора коэффициентов:

$$w = U\hat{w}, \quad \hat{w} = U^T w$$



# Решение задачи НК в новых признаках

Заменяем  $F$  на её приближение  $GU^T$ :

$$\|GU^T w - y\|^2 = \|G\hat{w} - y\|^2 \rightarrow \min_{\hat{w}}$$

Связь нового и старого вектора коэффициентов:

$$w = U\hat{w}, \quad \hat{w} = U^T w$$

Решение задачи наименьших квадратов относительно  $\hat{w}$  (единственное отличие –  $m$  слагаемых вместо  $n$ ):

$$\hat{w}^* = D^{-1}V^T y = \sum_{j=1}^m \frac{1}{\sqrt{\lambda_j}} u_j (v_j^T y)$$

$$G\hat{w}^* = VV^T y = \sum_{j=1}^m v_j (v_j^T y)$$

## Что почитать по этой лекции

- T. Hastie, R. Tibshirani "The Elements of Statistical Learning"  
Chapter 3

## На следующей лекции

- Bias-variance tradeoff
- Кривые обучения