

# Методы машинного обучения

## Линейные модели регрессии

Воронцов Константин Вячеславович

[www.MachineLearning.ru/wiki?title=User:Vokov](http://www.MachineLearning.ru/wiki?title=User:Vokov)

вопросы к лектору: [k.vorontsov@iai.msu.ru](mailto:k.vorontsov@iai.msu.ru)

материалы курса:

[github.com/MSU-ML-COURSE/ML-COURSE-25-26](https://github.com/MSU-ML-COURSE/ML-COURSE-25-26)

орг.вопросы по курсу: [ml.cmc@mail.ru](mailto:ml.cmc@mail.ru)

## 1 Многомерная линейная регрессия

- Метод наименьших квадратов
- Многомерная линейная регрессия
- Сингулярное разложение

## 2 Проблема мультиколлинеарности и регуляризация

- Проблема мультиколлинеарности
- $L_2$ -регуляризация: гребневая регрессия
- $L_1$ -регуляризация и отбор признаков

## 3 Метод главных компонент

- Постановка задачи и основная теорема
- Метод главных компонент для линейной регрессии
- Обобщения

## Оптимизационная постановка задач машинного обучения

$X$  — пространство *объектов*

$Y$  — множество *ответов* (предсказаний / оценок / прогнозов)

$y(x)$ ,  $y: X \rightarrow Y$  — неизвестная зависимость (target function)

**Дано:**

$\{x_1, \dots, x_\ell\} \subset X$  — *обучающая выборка* (training sample)

$a(x, w)$ ,  $a: X \times W \rightarrow Y$  — параметрическая модель зависимости

**Найти:**

$w \in W$  — вектор параметров модели такой, что  $a(x, w) \approx y(x)$

**Критерий** — минимум эмпирического риска:

$$\sum_{i=1}^{\ell} \mathcal{L}(w, x_i) \rightarrow \min_w \quad (\text{empirical risk minimization, ERM})$$

где  $\mathcal{L}(w, x)$  — *функция потерь* (loss function) — тем больше, чем сильнее  $a(x, w)$  отклоняется от правильного ответа  $y(x)$

## Многомерная линейная регрессия

$f_1(x), \dots, f_n(x)$  — числовые признаки объекта  $x$

$y_i = y(x_i) \in \mathbb{R}$  — «ответы учителя» на обучающих объектах

Модель многомерной линейной регрессии:

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

Матричные обозначения:

$$F_{\ell \times n} = \begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_\ell) & \dots & f_n(x_\ell) \end{pmatrix}, \quad y_{\ell \times 1} = \begin{pmatrix} y_1 \\ \dots \\ y_\ell \end{pmatrix}, \quad w_{n \times 1} = \begin{pmatrix} w_1 \\ \dots \\ w_n \end{pmatrix}.$$

Функционал квадрата ошибки:

$$Q(w, X^\ell) = \sum_{i=1}^{\ell} (a(x_i, w) - y_i)^2 = \|Fw - y\|^2 \rightarrow \min_w.$$

## Нормальная система уравнений

Необходимое условие минимума в матричном виде:

$$\frac{\partial Q(w)}{\partial w} = 2F^T(Fw - y) = 0,$$

откуда следует *нормальная система* задачи МНК:

$$F^T F w = F^T y,$$

где  $F^T F$  — матрица размера  $n \times n$  полного ранга  $n$ .

**Решение системы:**  $w^* = (F^T F)^{-1} F^T y = F^+ y$ .

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

где  $P_F = F F^+ = F(F^T F)^{-1} F^T$  — проекционная матрица.

## Геометрическая интерпретация МНК

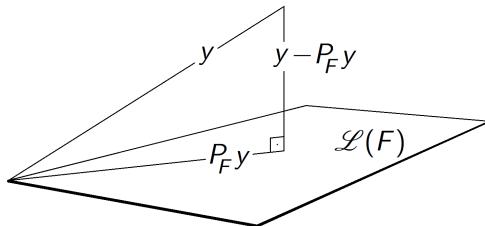
Линейная оболочка столбцов матрицы  $F = (f_1, \dots, f_n)$ ,  $f_j \in \mathbb{R}^\ell$ :

$$\mathcal{L}(F) = \left\{ \sum_{j=1}^n w_j f_j \mid w \in \mathbb{R}^n \right\}$$

$P_F = F(F^T F)^{-1} F^T$  — проекционная матрица

$P_F u$  — проекция вектора  $u \in \mathbb{R}^\ell$  на подпространство  $\mathcal{L}(F)$

$(I_\ell - P_F)u$  — проекция  $u$  на его ортогональное дополнение



МНК — это опускание перпендикуляра в  $\mathbb{R}^\ell$  из  $u$  на  $\mathcal{L}(F)$

## Сингулярное разложение

Произвольная  $\ell \times n$ -матрица представима в виде *сингулярного разложения* (singular value decomposition, SVD):

$$F = VDU^T.$$

**Основные свойства** сингулярного разложения:

- 1  $\ell \times n$ -матрица  $V = (v_1, \dots, v_n)$  ортогональна,  $V^T V = I_n$ , столбцы  $v_j$  — собственные векторы  $\ell \times \ell$ -матрицы  $FF^T$ ;
- 2  $n \times n$ -матрица  $U = (u_1, \dots, u_n)$  ортогональна,  $U^T U = I_n$ , столбцы  $u_j$  — собственные векторы  $n \times n$ -матрицы  $F^T F$ ;
- 3  $n \times n$ -матрица  $D$  диагональна,  $D = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ ,  $\lambda_j \geq 0$  — общие собственные значения матриц  $F^T F$  и  $FF^T$ .

## Решение МНК через сингулярное разложение

Псевдообратная  $F^+ = (F^T F)^{-1} F^T$ , вектор МНК-решения  $w^*$ ,  
 МНК-аппроксимация целевого вектора  $Fw^*$ :

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

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

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

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

Тождества:  $(AB)^{-1} = B^{-1}A^{-1}$ ,  $(AB)^T = B^T A^T$ ,  $\|w\|^2 = w^T w$



## Мультиколлинеарность и переобучение в линейных моделях

### Возможные причины переобучения:

- линейная зависимость (мультиколлинеарность) признаков:  
линейная модель:  $a(x, w) = \langle w, x \rangle$   
мультиколлинеарность:  $\exists u \in \mathbb{R}^n: \forall x \in X \quad \langle u, x \rangle = 0$   
неединственность решения:  $\forall \gamma \in \mathbb{R} \quad a(x, w) = \langle w + \gamma u, x \rangle$
- слишком мало объектов, слишком много признаков

### Проявления переобучения:

- слишком большие веса  $|w_j|$  разных знаков
- неустойчивость линейной модели  $\langle w, x \rangle$
- $Q(X^\ell) \ll Q(X^k)$

### Простой способ уменьшить переобучение:

- регуляризация  $\|w\| \rightarrow \min$  (сокращение весов, weight decay)

## Мультиколлинеарность и число обусловленности матрицы

Если  $\exists u \in \mathbb{R}^n$ :  $\sum_{n \times n} u \approx 0$ , то некоторые с.з.  $S$  близки к нулю

Число обусловленности  $n \times n$ -матрицы  $S$ :

$$\mu(S) = \|S\| \|S^{-1}\| = \frac{\max_{u: \|u\|=1} \|Su\|}{\min_{u: \|u\|=1} \|Su\|} = \frac{\lambda_{\max}}{\lambda_{\min}}$$

При умножении обратной матрицы на вектор,  $z = S^{-1}u$ , относительная погрешность усиливается в  $\mu(S)$  раз:

$$\frac{\|\delta z\|}{\|z\|} \leq \mu(S) \frac{\|\delta u\|}{\|u\|}$$

В нашем случае:  $S = F^T F$ ,  $u = F^T y$ ,  $w^* = S^{-1}u = (F^T F)^{-1} F^T y$ , погрешности измерения признаков  $f_j(x_i)$  и ответов  $y_i$  усиливаются в  $\mu(F^T F)$  раз!

## Стратегии устранения мультиколлинеарности

Если матрица  $S = F^T F$  плохо обусловлена, то:

- решение  $w^*$  неустойчиво и плохо интерпретируемо, содержит большие по модулю  $w_j^*$  разных знаков;
- $\|w^*\|$  велико;
- возникает переобучение:  
на обучении  $Q(w^*, X^\ell) = \|Fw^* - y\|^2$  мало;  
на контроле  $Q(w^*, X^k) = \|F'w^* - y'\|^2$  велико;

Стратегии устранения мультиколлинеарности и переобучения:

- 1 регуляризация:  $\|w\| \rightarrow \min$ ;
- 2 отбор признаков:  $f_1, \dots, f_n \rightarrow f_{j_1}, \dots, f_{j_m}, \quad m \ll n$ .
- 3 преобразование признаков:  $f_1, \dots, f_n \rightarrow g_1, \dots, g_m, \quad m \ll n$ ;

## Гребневая регрессия (ridge regression)

Штраф за увеличение  $L_2$ -нормы вектора весов  $\|w\|$ :

$$Q_\tau(w) = \|Fw - y\|^2 + \frac{\tau}{2} \|w\|^2,$$

где  $\tau$  — неотрицательный *параметр регуляризации*.

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

$$\begin{aligned} \frac{\partial Q_\tau(w)}{\partial w} &= 2F^\top(Fw - y) + 2\tau w = 0 \\ w_\tau^* &= (F^\top F + \tau I_n)^{-1} F^\top y. \end{aligned}$$

**Преимущество** сингулярного разложения:

можно подбирать параметр  $\tau$ , вычислив SVD только один раз.

## Регуляризованный МНК через сингулярное разложение

Вектор регуляризованного МНК-решения  $w_\tau^*$   
и МНК-аппроксимация целевого вектора  $Fw_\tau^*$ :

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

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

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

$Fw_\tau^* \neq Fw^*$ , но зато решение становится гораздо устойчивее.

## Выбор параметра регуляризации $\tau$

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

$$F'_{k \times n} = \begin{pmatrix} f_1(x'_1) & \dots & f_n(x'_1) \\ \dots & \dots & \dots \\ f_1(x'_k) & \dots & f_n(x'_k) \end{pmatrix}, \quad y'_{k \times 1} = \begin{pmatrix} y'_1 \\ \dots \\ y'_k \end{pmatrix}.$$

Вычисление функционала  $Q$  на контрольных данных  $T$  раз потребует  $O(kn^2 + knT)$  операций:

$$Q(w_\tau^*, X^k) = \|F' w_\tau^* - y'\|^2 = \left\| \underbrace{F' U}_{k \times n} \operatorname{diag}\left(\frac{\sqrt{\lambda_j}}{\lambda_j + \tau}\right) \underbrace{V^T y}_{n \times 1} - y' \right\|^2.$$

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

## Регуляризация сокращает «эффективную размерность»

Сжатие (shrinkage) или сокращение весов (weight decay):

$$\|w_{\tau}^*\|^2 = \sum_{j=1}^n \frac{\lambda_j}{(\lambda_j + \tau)^2} (v_j^T y)^2 < \|w^*\|^2 = \sum_{j=1}^n \frac{1}{\lambda_j} (v_j^T y)^2.$$

Почему говорят о сокращении эффективной размерности?

Роль размерности играет след проекционной матрицы:

$$\text{tr } F(F^T F)^{-1} F^T = \text{tr}(F^T F)^{-1} F^T F = \text{tr } I_n = n.$$

При использовании регуляризации:

$$\text{tr } F(F^T F + \tau I_n)^{-1} F^T = \text{tr } \text{diag} \left( \frac{\lambda_j}{\lambda_j + \tau} \right) = \sum_{j=1}^n \frac{\lambda_j}{\lambda_j + \tau} < n.$$

---

Тождество:  $\text{tr}(AB) = \text{tr}(BA)$

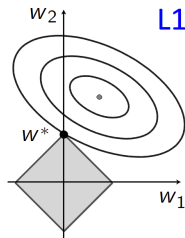
## Регуляризация по $L_1$ -норме для отбора признаков

LASSO — Least Absolute Shrinkage and Selection Operator

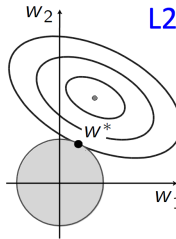
$$\|Fw - y\|^2 + \mu \sum_{j=1}^n |w_j| \rightarrow \min_w \iff \begin{cases} \|Fw - y\|^2 \rightarrow \min_w; \\ \sum_{j=1}^n |w_j| \leq \kappa; \end{cases}$$

LASSO ( $L_1$ ):

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



$L_1$



$L_2$

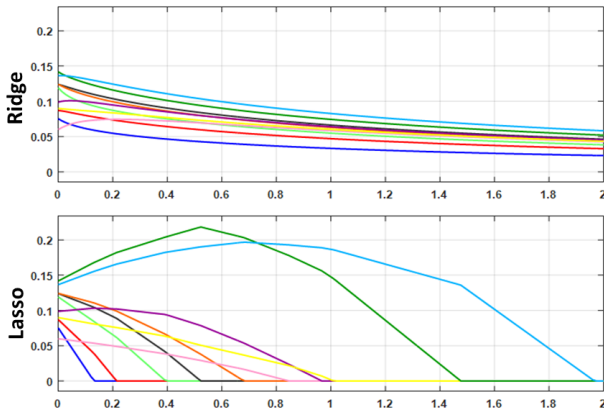
Ridge ( $L_2$ ):

$$\sum_{j=1}^n w_j^2 \leq \kappa$$



## Сравнение $L_2$ (Ridge) и $L_1$ (LASSO) регуляризации

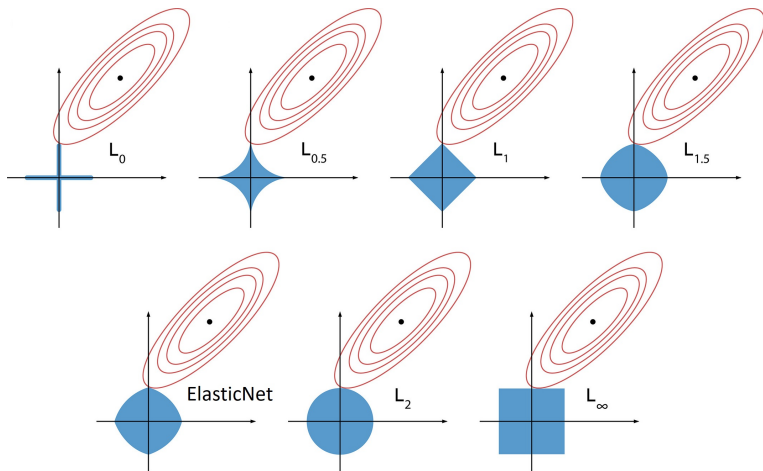
Типичный вид зависимости весов  $w_j$  от селективности  $\mu$



В LASSO с увеличением  $\mu$  усиливается отбор признаков

## Геометрическая интерпретация отбора признаков

Сравнение регуляризаторов по различным  $L_p$ -нормам:



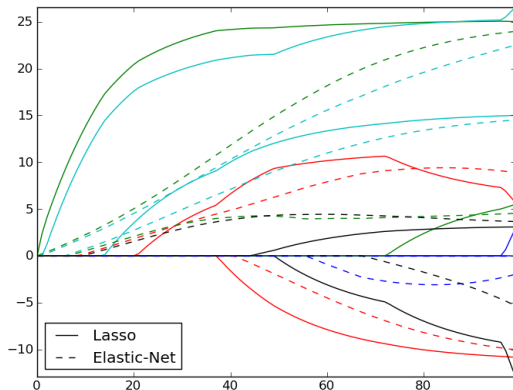
## Двойной регуляризатор $L_1 + L_2$ (Elastic Net)

$$\frac{1}{2} \|Fw - y\|^2 + \mu \sum_{j=1}^n |w_j| + \frac{\tau}{2} \sum_{j=1}^n w_j^2 \rightarrow \min_w$$

- ⊕ Параметр селективности  $\mu$  управляет отбором признаков: чем больше  $\mu$ , тем меньше признаков останется
- ⊕ Есть эффект группировки (grouping effect): значимые зависимые признаки отбираются вместе и имеют примерно равные веса  $w_j$
- ⊖ Приходится подбирать два параметра регуляризации  $\mu$ ,  $\tau$  (есть специальные методы — regularization path)
- ⊖ Шумовые признаки также группируются вместе; по мере увеличения  $\mu$  группы значимых признаков могут отбрасываться, когда ещё не все шумовые отброшены

## Двойной регуляризатор $L_1 + L_2$ (Elastic Net)

Elastic Net менее жёстко отбирает признаки, чем LASSO.  
Зависимости весов  $w_j$  от коэффициента  $\log \frac{1}{\mu}$ :

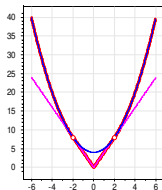


[scikit-learn.org/0.5/auto\\_examples/glm/plot\\_lasso\\_coordinate\\_descent\\_path.html](http://scikit-learn.org/0.5/auto_examples/glm/plot_lasso_coordinate_descent_path.html)

## Отбор опорных признаков (Support Features Machine, SFM)

Гладкая склейка  $L_1$  и  $L_2$  регуляризаторов:

$$\frac{1}{2} \|Fw - y\|^2 + \sum_{j=1}^n R_{\mu}(w_j) \rightarrow \min_w$$
$$R_{\mu}(w_j) = \begin{cases} 2\mu|w_j|, & |w_j| \leq \mu \\ \mu^2 + w_j^2, & |w_j| \geq \mu \end{cases}$$



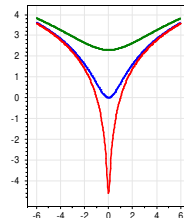
- ⊕ Только один параметр регуляризации  $\mu$
- ⊕ Отбор признаков с *параметром селективности*  $\mu$
- ⊕ *Эффект группировки*: значимые зависимые признаки ( $|w_j| > \mu$ ) входят в решение совместно (как в Elastic Net)
- ⊕ Шумовые признаки ( $|w_j| < \mu$ ) не группируются и подавляются независимо друг от друга (как в LASSO)

Tatarchuk A., Urlov E., Mottl V., Windridge D. A support kernel machine for supervised selective combining of diverse pattern-recognition modalities. 2010.

## Отбор релевантных признаков (Relevance Feature Machine, RFM)

Невыпуклый регуляризатор:

$$\frac{1}{2} \|Fw - y\|^2 + \sum_{j=1}^n \ln(w_j^2 + \frac{1}{\mu}) \rightarrow \min_w$$
$$R(w) = \ln(w^2 + \frac{1}{\mu}) \text{ при } \mu = 0.1, 1, 100$$



- ⊕ Только один параметр регуляризации  $\mu$
- ⊕ Отбор признаков с параметром *селективности*  $\mu$
- ⊕ Есть эффект группировки
- ⊕ Лучше отбирает набор значимых признаков, когда они только совместно обеспечивают хорошее решение

---

*Tatarchuk A., Mottl V., Eliseyev A., Windridge D.* Selectivity supervision in combining pattern recognition modalities by feature- and kernel-selective Support Vector Machines. 2008.

## Метод главных компонент (Principal Component Analysis, PCA)

$f_1(x), \dots, f_n(x)$  — исходные числовые признаки;

$g_1(x), \dots, g_m(x)$  — новые числовые признаки,  $m \leq n$ ;

**Требование:** старые признаки  $f_j(x)$  должны линейно восстанавливаться по новым признакам  $g_s(x)$ :

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

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

$$\sum_{i=1}^{\ell} \sum_{j=1}^n (\hat{f}_j(x_i) - f_j(x_i))^2 \rightarrow \min_{\{g_s(x_i)\}, \{u_{js}\}}$$

Задача преобразования признаков (feature transformation)

— это **задача обучения без учителя**, тут нет ответов  $y_i$

## Матричные обозначения

Матрицы «объекты–признаки», старая и новая:

$$F_{\ell \times n} = \begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_\ell) & \dots & f_n(x_\ell) \end{pmatrix}; \quad G_{\ell \times m} = \begin{pmatrix} g_1(x_1) & \dots & g_m(x_1) \\ \dots & \dots & \dots \\ g_1(x_\ell) & \dots & g_m(x_\ell) \end{pmatrix}.$$

Матрица линейного преобразования новых признаков в старые:

$$U_{n \times m} = \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \dots & \dots & \dots \\ u_{n1} & \dots & u_{nm} \end{pmatrix}; \quad \hat{F} = GU^T \overset{\text{хотим}}{\approx} F.$$

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

$$\sum_{i=1}^{\ell} \sum_{j=1}^n (\hat{f}_j(x_i) - f_j(x_i))^2 = \|GU^T - F\|^2 \rightarrow \min_{G, U},$$



## Основная теорема метода главных компонент

### Теорема

Если  $m \leq \text{rk } F$ , то минимум  $\|GU^T - F\|^2$  достигается, когда столбцы  $U$  — это с.в. матрицы  $F^T F$ , соответствующие  $m$  максимальным с.з.  $\lambda_1, \dots, \lambda_m$ , а матрица  $G = FU$ .

При этом:

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

Тождество:  $\|F\|^2 = \text{tr}(F^T F) = \sum_{j=1}^n \lambda_j$

## Связь с сингулярным разложением

Произвольная  $\ell \times n$ -матрица  $F$  представима в виде SVD:

$$F = VDU^T; \quad U^T U = I_m; \quad V^T V = I_m; \quad D = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$$

Если взять  $m = n$ , то:

- 1  $\|GU^T - F\|^2 = 0$
- 2 представление  $\hat{F} = GU^T = F$  точное и совпадает с SVD, если положить  $G = V\sqrt{\Lambda}$ ,  $D = \sqrt{\Lambda}$ :

$$F = GU^T = V\sqrt{\Lambda}U^T;$$

- 3 линейное преобразование  $U$  работает в обе стороны:

$$F = GU^T; \quad G = FU.$$

$G^T G = \Lambda$  — новые признаки некоррелированы,  
 $U$  — декоррелирующее преобразование Карунена–Лоэва

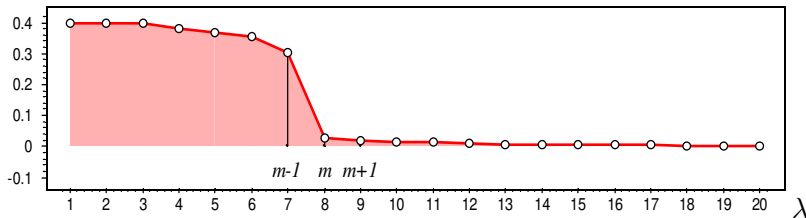
## Эффективная размерность выборки

Упорядочим с.з.  $F^T F$  по убыванию:  $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ .

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

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

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



## Решение задачи НК в новых признаках

Заменяем  $F_{\ell \cdot n}$  на её приближение  $G_{\ell \cdot m} \cdot U_{m \cdot n}^T$ , предполагая  $m \leq n$ :

$$\|G \underbrace{U^T w}_{z} - y\|^2 = \|Gz - y\|^2 \rightarrow \min_z.$$

Связь нового и старого вектора коэффициентов:

$$z = U^T w; \quad w = Uz.$$

Решение задачи наименьших квадратов относительно  $z$ :

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

$$Gz^* = V V^T y = \sum_{j=1}^m v_j (v_j^T y);$$

Единственное отличие от прежнего  $w^*$  —  $m$  слагаемых вместо  $n$

## Спектральный метод наименьших квадратов

1. Построить SVD-разложение, упорядочить  $\lambda_1 \geq \dots \geq \lambda_n$
2. Отделить  $n - m$  наименьших с.з. от нуля:  $\lambda'_j := \lambda_j + \delta_j$

Частные случаи:

- $\lambda'_j := \lambda_j + \tau$  — гребневая регрессия
- $\lambda'_j := \lambda_j + \infty[j > m]$  — метод главных компонент
- $\lambda'_j := \lambda_j + \tau[j > m]$  — нечто промежуточное

3. Применить формулы SVD для модификации МНК-решения:

$$w^* = \sum_{j=1}^n \frac{1}{\sqrt{\lambda_j}} u_j(v_j^T y) \quad \longrightarrow \quad w^* = \sum_{j=1}^n \frac{\sqrt{\lambda_j}}{\lambda'_j} u_j(v_j^T y)$$

$$Fw^* = \sum_{j=1}^n v_j(v_j^T y) \quad \longrightarrow \quad Fw^* = \sum_{j=1}^n \frac{\lambda_j}{\lambda'_j} v_j(v_j^T y)$$

- Многомерная линейная регрессия  
— через *сингулярное разложение*
- Три приёма против мультиколлинеарности и переобучения  
— регуляризация, отбор и преобразование признаков
- $L_2$ -регуляризация, она же гребневая регрессия  
— тоже через *сингулярное разложение*
- $L_1$ -регуляризация (LASSO) и др. негладкие регуляризаторы  
— регулируемый отбор признаков
- Метод главных компонент — задача матричного разложения  
— снова через *сингулярное разложение*
- Другие методы матричных разложений и их приложения  
— в следующем семестре