Методы детектирования аномалий.
Лекция 5: Линейные модели

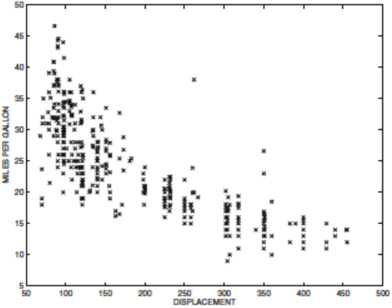
Иван Шанин
ivan.shanin@gmail.com

ИПИ РАН

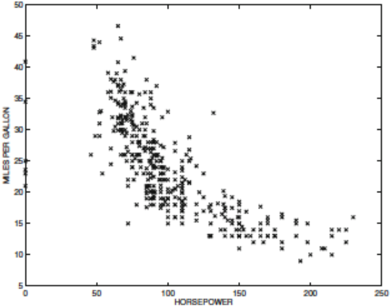
18.03.2019

Регрессионные модели

- **Предположение:** датасет на самом деле лежит в линейном подпространстве меньшей размерности
- Подобное предположение осмысленно при наличии ярко выраженной мультиколлинарности в задаче
- А также при наличии неинформативных признаков



(a) View 1 (*Autompg*)



(b) View 2 (*Autompg*)

Линейные регрессионные модели

- Выразим признак y через зависимые переменные:

$$y = \sum_{i=1}^{d} w_i \cdot x_i + w_{d+1}$$

- Найдем параметры $w_1 \dots w_{d+1}$ при помощи обучающей выборки $\overline{X_1} \dots \overline{X_N}$
- Тогда для каждого объекта $\overline{X_j}$ вычисляется **ошибка** ϵ_j смоделированного признака y:

$$y_j = \sum_{i=1}^{d} w_i \cdot x_{ij} + w_{d+1} + \epsilon_j$$

- Суммарный квадрат ошибки $\sum_{j=1}^{N} \epsilon_j^2$ можно использовать как функционал качества задачи регрессии (регрессия методом наименьших квадратов)
- Значение ошибки на конкретном объекте является показателем аномальности объекта

Матричная постановка задачи

- составим матрицу D размерности $N \times d + 1$ из строк $(x_{j1}, \ldots, x_{jd}, 1)$
- также составим вектор весов $\overline{W}$ и вектор значений признака $\overline{y}$
- задачу нахождения вектора $\overline{W}$ можно записать следующей системой уравнений

$$\overline{y} \approx D\overline{W}^T$$

$$\|D\overline{W}^T - \overline{y}\|^2 \to \min \quad \Rightarrow \quad 2D^T D\overline{W}^T - 2D^T\overline{y} = 0 \quad \Rightarrow$$

$$\overline{W}^T = (D^T D)^{-1} D^T \overline{y}$$

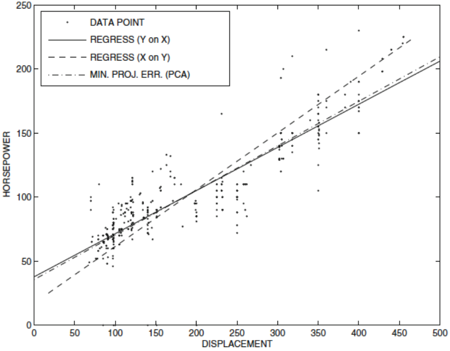
- Таким образом, для нахождения оптимального набора параметров $\overline{W}$ необходимо обратить матрицу $D^T D$

Гребневая регуляризация

- Для борьбы с недоопределенностью используется регуляризатор:

$$J = \|D\overline{W}^T - \overline{y}\|^2 + \alpha\|W\|^2 \to \min_{\overline{W}}$$

- $2(D^T D + \alpha I)\overline{W}^T - 2D^T\overline{y} = 0 \quad \Rightarrow \quad \overline{W}^T = (D^T D)^{-1} D^T \overline{y}$
- Полученный показатель аномальности будет зависеть от выбора моделируемого признака



Principal Component Analysis: постановка задачи

- Рассмотрим выборку

$$F_{L\times N} = \begin{bmatrix} f_1(x_1) & \dots & f_N(x_1) \\ \dots & \dots & \dots \\ f_1(x_L) & \dots & f_N(x_L) \end{bmatrix} = \begin{bmatrix} x_1 \\ \dots \\ x_L \end{bmatrix}$$

- Обозначим через $z_i = g_1(x_i), \dots, g_M(x_i)$ признаковые описания тех же объектов в новом пространстве размерности $M < N$

$$G_{L\times M} = \begin{bmatrix} g_1(x_1) & \dots & g_M(x_1) \\ \dots & \dots & \dots \\ g_1(x_L) & \dots & g_M(x_L) \end{bmatrix} = \begin{bmatrix} x_1 \\ \dots \\ x_L \end{bmatrix}$$

- Определим матрицу линейного преобразования $U = (u_{js})_{N\times M}$:

$$\hat{f}_j(x) = \sum_{s=1}^{M} g_s(x) u_{js}$$

Метод главных компонент

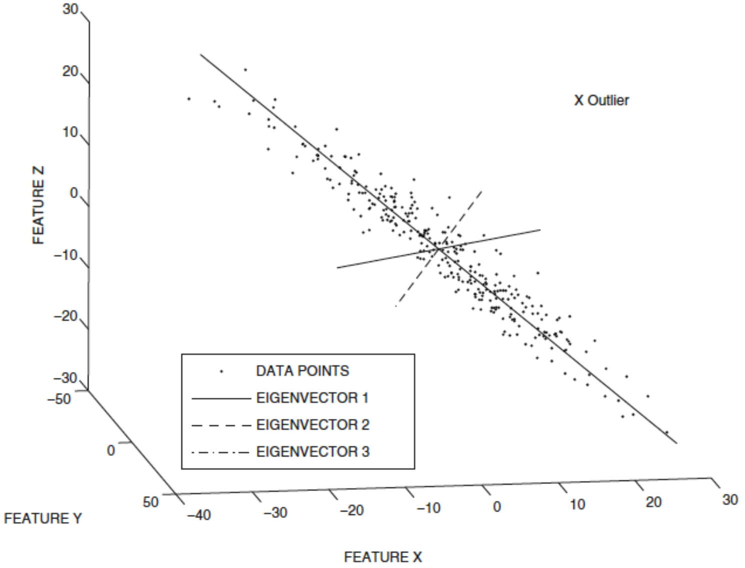
- Будем искать одновременно матрицы G и U, минимизируя суммарную невязку смоделированных признаков

$$\Delta^2(G,U) = \sum_{i=1}^{L} \|\hat{x}_i - x_i\|^2 = \sum_{i=1}^{L} \|z_i U^T - x_i\|^2 = \|GU^T - F\|^2 \to \min_{G,U}$$

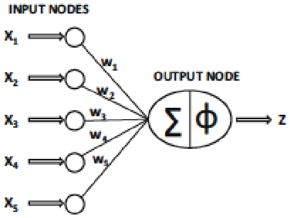
Теорема (PCA)

Если $m \leq \operatorname{rank} F$, то минимум $\Delta^2(G,U)$ достигается, когда столбцы матрицы U есть собственные векторы $F^T F$, соответствующие m максимальным собственным значениям. При этом $G = FU$, матрицы U и G ортогональны

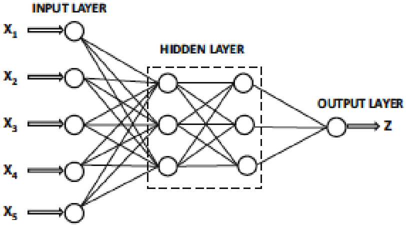
Метод главных компонент



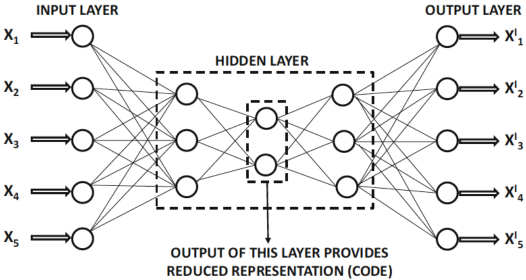
Автоэнкодеры



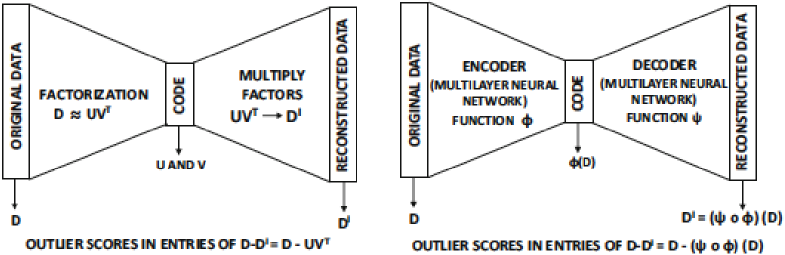
(a) Perceptron



(b) Multilayer



Автоэнкодеры и матричная факторизация



(a) Matrix factorization (b) Autoencoder