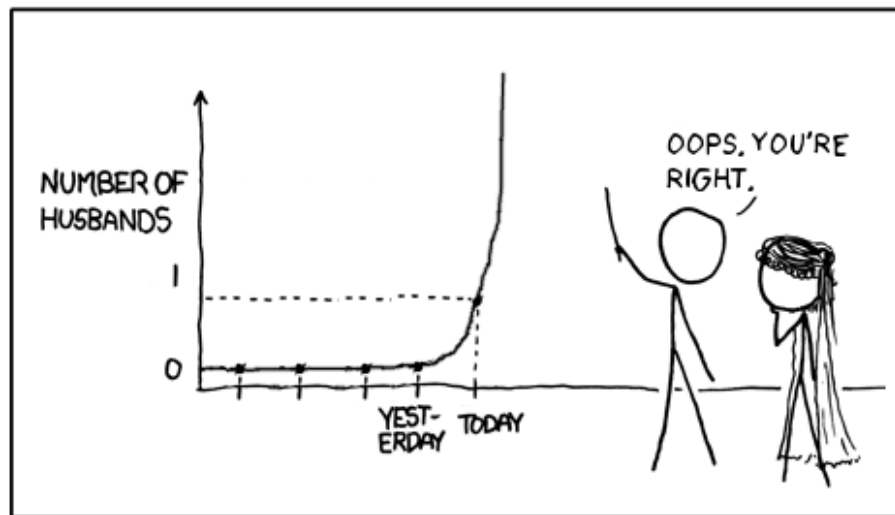
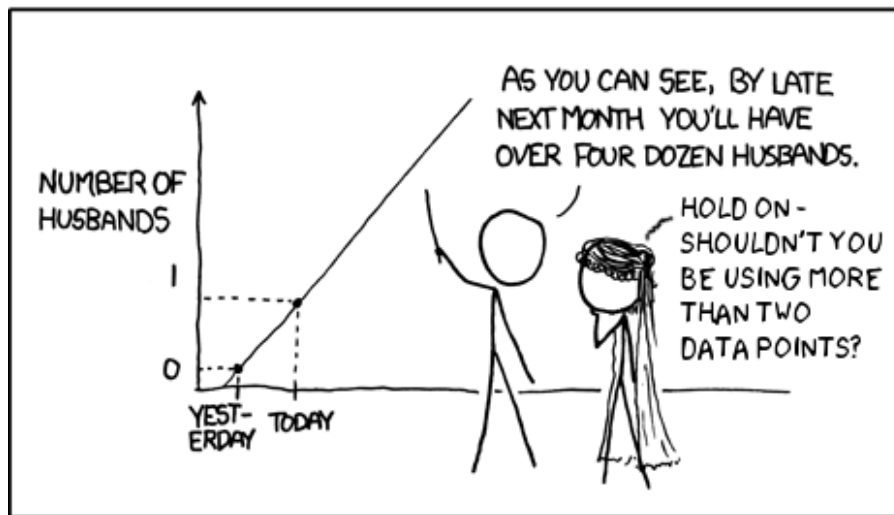


Эволюция

или как превратить линейную модель в
ИСКУССТВО

Воспоминания о линейных моделях

MY HOBBY: EXTRAPOLATING



Воспоминания о линейных моделях

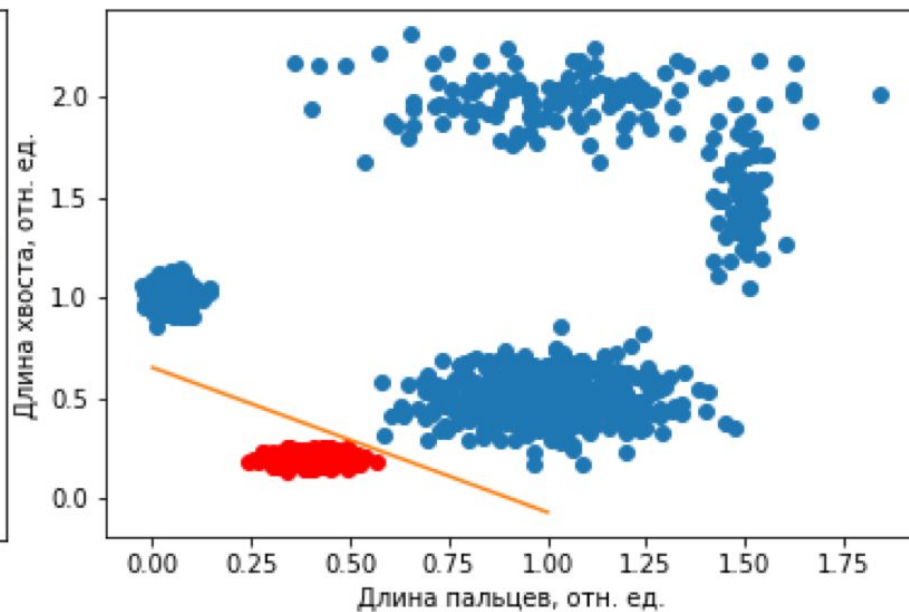
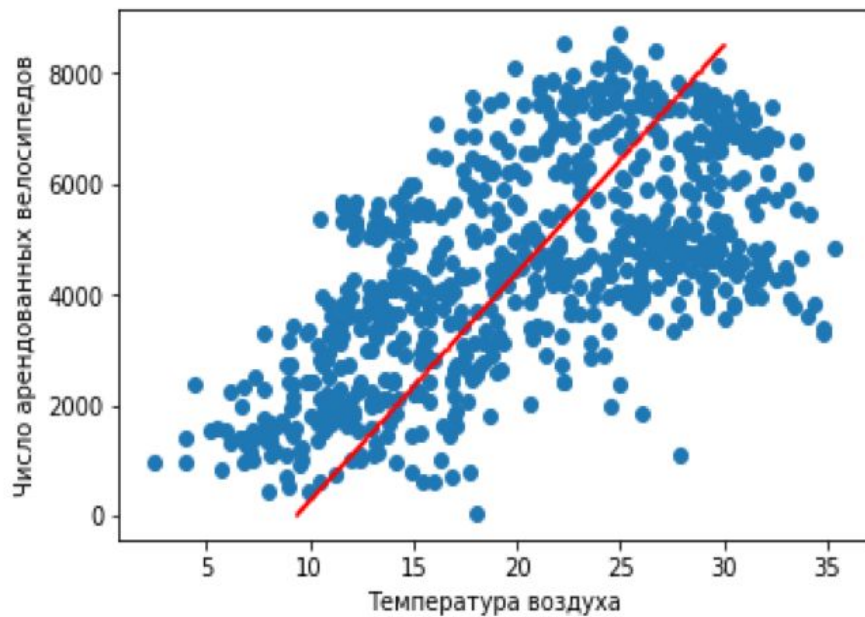
Линейная модель представима в виде:

$$y = w_0 + \sum_{i=1}^m w_i x_i$$

Или более строго:

$$y = \sum_{i=0}^m w_i x_i = \vec{w}^T \vec{x}$$

Воспоминания о линейных моделях



Регуляризаторы

Проблема большого количества признаков – наличие зависимых друг от друга признаков, ведущее к неустойчивости модели, и переобучение модели под обучающую выборку.

Решение – добавление к штрафу за большое отклонение от требуемого значения (квадратичному отклонению) штрафа за большие веса. Весовой штраф может быть выражен как модулями, так и квадратами модулей весов признаков.

Регуляризаторы

$$R(w) = ||w||_2^2 = \sum_{i=1}^d w_i^2$$
$$R(w) = ||w||_1 = \sum_{i=1}^d |w_i|$$

Отбор признаков

Рассмотрим отрезок $[0,1]$ в одномерном пространстве:

- в отрезке $[0.1,0.99]$ лежат 98% всех точек

Рассмотрим отрезок (куб) $[0,1]^3$ в трёхмерном пространстве:

- в кубе $[0.1,0.99]^3$ лежат уже ~94% всех точек

Рассмотрим отрезок $[0,1]^{1000}$ в тысячеммерном пространстве:

- в кубе $[0.1,0.99]^{1000}$ лежат всего $\sim 1.6 \times 10^{-7}\%$ всех точек

Нормализация данных

Необходима при различных масштабах признаков.

