



ТЕХНОСФЕРА

Лекция 9 Линейные модели вероятностная перспектива

Николай Анохин

18 апреля 2016 г.

План занятия

Линейная регрессия

Логистическая регрессия

Постановка задачи

Дано. Признаковые описания N объектов $\mathbf{x} = (x_1, \dots, x_m) \in \mathcal{X}$, образующие тренировочный набор данных X , и значения целевой переменной $y = f(\mathbf{x}) \in \mathcal{Y}$ для каждого объекта из X .

Найти. Для семейства параметрических функций

$$H = \{h(\mathbf{x}, \theta) = y : \mathcal{X} \times \Theta \rightarrow \mathcal{Y}\},$$

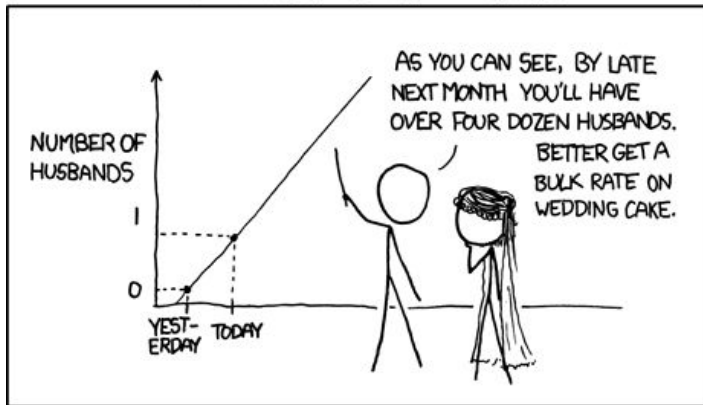
найти значение вектора параметров θ^* , такое что $h^*(\mathbf{x}) = h(\mathbf{x}, \theta^*)$ наилучшим образом приближает целевую функцию.

$\mathcal{Y} \in \{C_1, C_2, \dots, C_K\}$ – задача классификации

$\mathcal{Y} \in [a, b] \subset \mathcal{R}$ – задача регрессии

Линейная регрессия

MY HOBBY: EXTRAPOLATING



Модель

$$y = h(\mathbf{x}, \theta) + \epsilon,$$

где ϵ – гауссовский шум

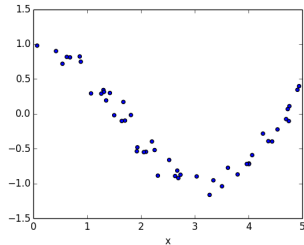
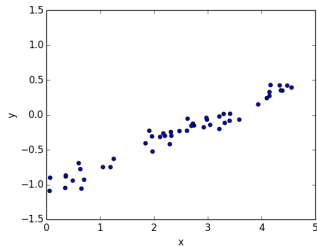
$$p(\epsilon) = \mathcal{N}(\epsilon|0, \beta^{-1}),$$

откуда

$$p(y|\mathbf{x}, \theta, \beta) = \mathcal{N}(y|h(\mathbf{x}, \theta), \beta^{-1}).$$

Предсказание

$$E[y|\mathbf{x}] = \int y p(y|\mathbf{x}) dy = h(\mathbf{x}, \theta).$$



Линейная модель

простейшая модель

$$h(\mathbf{x}, \mathbf{w}) = w_0 + w_1x_1 + \dots + w_Mx_M = \sum_{j=0}^M w_jx_j$$

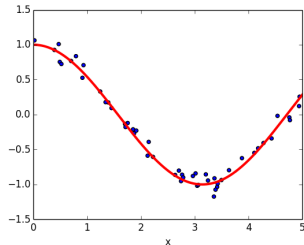
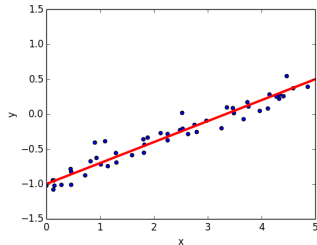
улучшенная модель

$$h(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^M w_j\phi_j(\mathbf{x}) = \mathbf{w}^T\phi(\mathbf{x}),$$

$\phi_j(\mathbf{x})$ – базисные функции, $\phi_0(\mathbf{x}) = 1$

примеры

$$\varphi_j(x) = x^j, \quad \varphi_j(x) = \exp\left\{-\frac{(x - \mu_j)^2}{2s^2}\right\}$$



ML – функция правдоподобия

Дана обучающая выборка $\mathcal{D} = (X, Y)$ из N объектов (\mathbf{x}_n, y_n)

Функция правдоподобия

$$\begin{aligned}\log p(Y|X, \mathbf{w}, \beta) &= \sum_{n=1}^N \log \mathcal{N}(y_n | \mathbf{w}^T \phi(\mathbf{x}_n), \beta^{-1}) = \\ &= \frac{N}{2} \log \beta - \frac{N}{2} \log 2\pi - \frac{\beta}{2} \sum_{n=1}^N \{y_n - \mathbf{w}^T \phi(\mathbf{x}_n)\}^2 \rightarrow \max_{\mathbf{w}, \beta}\end{aligned}$$

Квадратичная функция потерь

$$E_D(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y_n - \mathbf{w}^T \phi(\mathbf{x}_n)\}^2 \rightarrow \min_{\mathbf{w}}$$

ML – решение

$$\log p(Y|X, \mathbf{w}, \beta) = \frac{N}{2} \log \beta - \frac{N}{2} \log 2\pi - \frac{\beta}{2} \sum_{n=1}^N \{y_n - \mathbf{w}^T \phi(\mathbf{x}_n)\}^2 \rightarrow \max_{\mathbf{w}, \beta}$$

Градиент

$$\beta \sum_{n=1}^N \{y_n - \mathbf{w}^T \phi(\mathbf{x}_n)\} \phi(\mathbf{x}_n)^T = 0$$

Решение

$$\mathbf{w}_{ML} = \Phi^\dagger Y = (\Phi^T \Phi)^{-1} \Phi^T Y, \quad \frac{1}{\beta_{ML}} = \frac{1}{N} \sum_{n=1}^N \{y_n - \mathbf{w}_{ML}^T \phi(\mathbf{x}_n)\}^2,$$

где

$$\Phi = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \dots & \phi_M(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \dots & \phi_M(\mathbf{x}_2) \\ \dots & \dots & \dots \\ \phi_0(\mathbf{x}_N) & \dots & \phi_M(\mathbf{x}_N) \end{pmatrix}$$

Регуляризация

Функция потерь

$$E(\mathbf{w}, \lambda) = E_D(\mathbf{w}) + \lambda E_W(\mathbf{w}),$$

где (как и раньше)

$$E_D(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y_n - \mathbf{w}^T \phi(\mathbf{x}_n)\}^2 \rightarrow \min_{\mathbf{w}},$$

плюс регуляризация

$$E_W(\mathbf{w}) = E_q(\mathbf{w}) = \sum_{j=1}^M |\mathbf{w}_j|^q$$

Зоопарк

- ▶ $q = 1$ – Lasso
- ▶ $q = 2$ – Ridge (байесовский вывод: $p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|0, \alpha^{-1}\mathbf{I})$)
- ▶ $E_W(\mathbf{w}) = \rho E_1(\mathbf{w}) + (1 - \rho)E_2(\mathbf{w})$ – Elastic Net

Робастная регрессия

Проблема: линейная регрессия чувствительна к шуму из-за того, что плотность нормального распределение очень быстро убывает

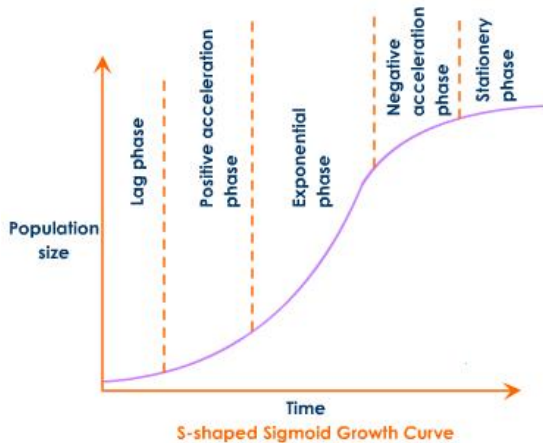
Решения:

- ▶ MAE вместо MSE
- ▶ LTS: Least Trimmed Sum of Squares ¹
- ▶ RANSAC ²
- ▶ t-Distribution вместо Гаусса

¹Least trimmed squares (wikipedia)

²RANSAC (wikipedia)

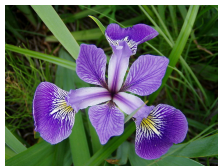
Логистическая регрессия



Ирисы Фишера



Setosa



Versicolor

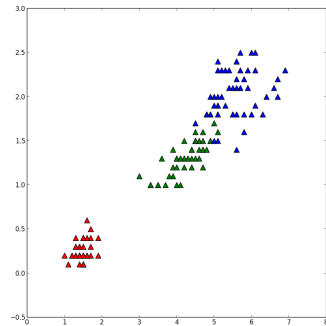
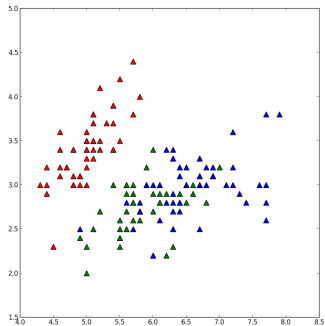


Virginica

Задача

Определить вид ириса на основании длины чашелистика, ширины чашелистика, длины лепестка и ширины лепестка.

Ирисы Фишера



Многомерное нормальное распределение

$$\mathcal{N}(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu) \right\}$$

Параметры

D -мерный вектор средних

$D \times D$ -мерная матрица ковариации

$$\mu = \int \mathbf{x} p(\mathbf{x}) d\mathbf{x}$$

$$\Sigma = E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T]$$

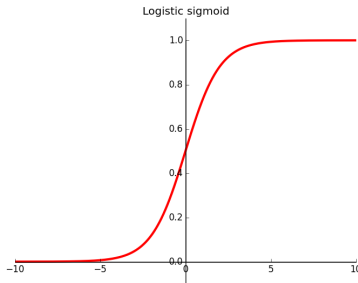
Генеративная модель

Рассматриваем 2 класса

$$p(y_1|x) = \frac{p(x|y_1)p(y_1)}{p(x|y_1)p(y_1) + p(x|y_2)p(y_2)} = \frac{1}{1 + e^{-a}} = \sigma(a)$$

$$a = \ln \frac{p(x|y_1)p(y_1)}{p(x|y_2)p(y_2)}$$

$\sigma(a)$ – сигмоид-функция, $a = \ln(\sigma/(1 - \sigma))$



Случай нормальных распределений

Пусть

$$p(\mathbf{x}|y_k) = \mathcal{N}(\mathbf{x}|\mu_k, \mathbf{\Sigma}),$$

тогда

$$p(y_1|\mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x} + w_0),$$

где

$$\mathbf{w} = \mathbf{\Sigma}^{-1}(\mu_1 - \mu_2)$$

$$w_0 = -\frac{1}{2}\mu_1^T \mathbf{\Sigma}^{-1} \mu_1 + \frac{1}{2}\mu_2^T \mathbf{\Sigma}^{-1} \mu_2 + \ln \frac{p(y_1)}{p(y_2)}$$

Аналогичный результат для любых распределений из экспоненциального семейства

Maximum Likelihood

$$p(y_1, \mathbf{x}) = p(y_1)p(\mathbf{x}|y_1) = \pi \mathcal{N}(\mathbf{x}|\mu_1, \mathbf{\Sigma})$$

$$p(y_2, \mathbf{x}) = p(y_2)p(\mathbf{x}|y_2) = (1 - \pi) \mathcal{N}(\mathbf{x}|\mu_2, \mathbf{\Sigma})$$

Функция правдоподобия

$$p(Y, X|\pi, \mu_1, \mu_2, \mathbf{\Sigma}) = \prod_{n=1}^N [\pi \mathcal{N}(\mathbf{x}|\mu_1, \mathbf{\Sigma})]^{y_n} [(1 - \pi) \mathcal{N}(\mathbf{x}|\mu_2, \mathbf{\Sigma})]^{1-y_n}$$

Максимизируя $\log p(Y, X|\pi, \mu_1, \mu_2, \mathbf{\Sigma})$, имеем

$$\pi = \frac{1}{N} \sum_{n=1}^N y_n = \frac{N_1}{N_1 + N_2},$$

$$\mu_1 = \frac{1}{N_1} \sum_{n=1}^N y_n \mathbf{x}_n, \quad \mu_2 = \frac{1}{N_2} \sum_{n=1}^N (1 - y_n) \mathbf{x}_n,$$

аналогично для $\mathbf{\Sigma}$

Логистическая регрессия

Дано.

$$\mathcal{D} = \{\phi_n = \phi(\mathbf{x}_n), y_n\}, y_n \in \{0, 1\}, n = 1 \dots N$$

Модель.

$$p(y = 1|\phi) = \sigma(\mathbf{w}^\top \phi)$$

функция правдоподобия (кросс-энтропия)

$$\begin{aligned} l(\mathbf{w}) &= \log \left[\prod_{n=1}^N p^{y_n}(y = 1|\phi_n)(1 - p(y = 1|\phi_n))^{1-y_n} \right] = \\ &= \sum_{n=1}^N y_n \log p(y = 1|\phi_n) + (1 - y_n) \log(1 - p(y = 1|\phi_n)) = -J_c(\mathbf{w}) \rightarrow \max_{\mathbf{w}} \end{aligned}$$

Градиент

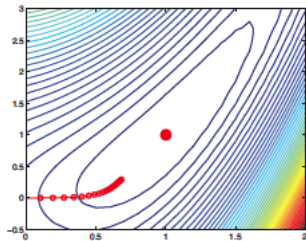
$$\nabla J_c(\mathbf{w}) = \sum_{n=1}^N (p(y=1|\phi_n) - y_n) \phi_n$$

Гессиан

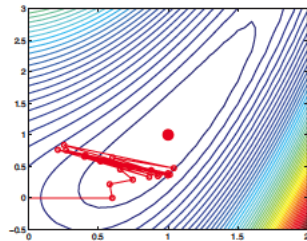
$$\nabla^2 J_c(\mathbf{w}) = \sum_{n=1}^N p(y=1|\phi_n)(1 - p(y=1|\phi_n)) \phi_n \phi_n^T$$

Градиентный спуск

```
1 function gd(grad, a0, epsilon):  
2     initialise eta(k)  
3     k = 0  
4     a = a0  
5     do:  
6         k = k + 1  
7         a = a - eta(k) grad(a)  
8     until eta(k) grad(a) < epsilon  
9     return a
```



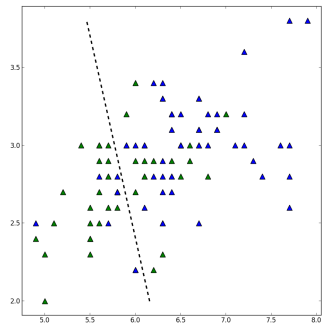
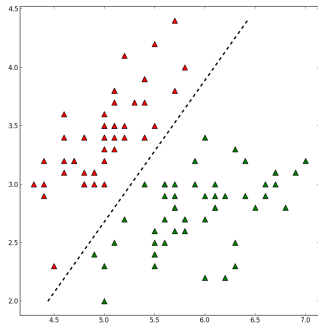
(a)



(b)

Добавление момента: $\mathbf{a}_{k+1} = \mathbf{a}_k - \eta_k \nabla J(\mathbf{a}_k) + \mu_k(\mathbf{a}_k - \mathbf{a}_{k-1})$

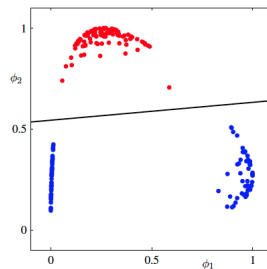
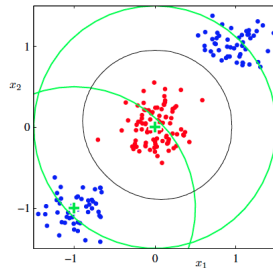
Логистическая регрессия: результаты



Обобщенная линейная модель / GLM

$$y(\mathbf{x}, \mathbf{w}) \sim pdf \left[f(\mathbf{w}^\top \phi(\mathbf{x})) \right],$$

- ▶ $\phi_n(\mathbf{x})$ – базисные функции
- ▶ $f(a)$ – функция активации
- ▶ pdf – распределение из экспоненциального семейства



Distribution	Support of distribution	Typical uses	Link name	Link function	Mean function
Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbf{X}\beta = \mu$	$\mu = \mathbf{X}\beta$
Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Inverse	$\mathbf{X}\beta = -\mu^{-1}$	$\mu = -(\mathbf{X}\beta)^{-1}$
Gamma					
Inverse Gaussian	real: $(0, +\infty)$		Inverse squared	$\mathbf{X}\beta = -\mu^{-2}$	$\mu = (-\mathbf{X}\beta)^{-1/2}$
Poisson	integer: $0, 1, 2, \dots$	count of occurrences in fixed amount of time/space	Log	$\mathbf{X}\beta = \ln(\mu)$	$\mu = \exp(\mathbf{X}\beta)$
Bernoulli	integer: $\{0, 1\}$	outcome of single yes/no occurrence	Logit	$\mathbf{X}\beta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(\mathbf{X}\beta)}{1 + \exp(\mathbf{X}\beta)} = \frac{1}{1 + \exp(-\mathbf{X}\beta)}$
Binomial	integer: $0, 1, \dots, N$	count of # of "yes" occurrences out of N yes/no occurrences			
Categorical	integer: $[0, K)$	outcome of single K-way occurrence			
	K-vector of integer: $[0, 1]$, where exactly one element in the vector has the value 1				
Multinomial	K-vector of integer: $[0, N]$	count of occurrences of different types (1 .. K) out of N total K-way occurrences			

Источник: Wikipedia

Мультикласс классификация

- ▶ one-vs-rest

Строим K моделей, каждая соответствует одному классу

- ▶ one-vs-one

Строим $K(K - 1)/2$ моделей, каждая соответствует паре классов

Вопросы

