ТЕХНОСФЕРА

Лекция 6
Линейные модели для классификации и регрессии

Николай Анохин

28 октября 2014 г.

План занятия

Линейная регрессия

Логистическая регрессия

Обобщенные линейные модели

Постановка задачи

Пусть дан набор объектов $\mathcal{D} = \{(\mathbf{x}_i, y_i)\},\ \mathbf{x}_i \in \mathcal{X},\ y_i \in \mathcal{Y},\ i \in 1, \dots, N$, полученный из неизвестной закономерности $y = f(\mathbf{x})$. Необходимо выбрать из семейства параметрических функций

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

такую $h^*(\mathbf{x}) = h(\mathbf{x}, \theta^*)$, которая наиболее точно апроксимирует $f(\mathbf{x})$.

Задачи

- Регрессия: $\mathcal{Y} = [a, b] \subset \mathbb{R}$
- Классификация: $|\mathcal{Y}| < C$

Линейная регрессия

Модель

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

где ϵ – гауссовский шум

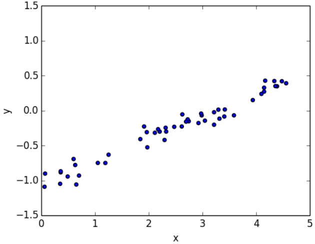
$$p(\epsilon) = \mathcal{N}(\epsilon | 0, \beta^{-1}),$$

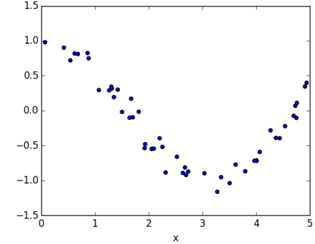
откуда

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

Предсказание

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





Линейная модель

простейшая модель

$$h(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + \ldots + w_M x_M = \sum_{j=0}^{M} w_j x_j$$

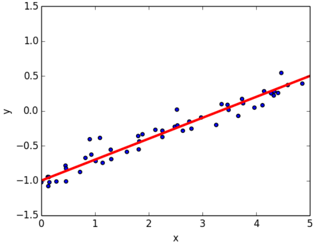
улучшенная модель

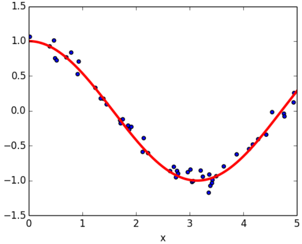
$$h(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M} w_j \phi_j(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}),$$

$\phi_j(\mathbf{x})$ – 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

$$\log p(Y|X, \mathbf{w}, \beta) = \sum_{n=1}^{N} \log \mathcal{N}(y|\mathbf{w}^T \phi(\mathbf{x_n}), \beta^{-1}) =$$

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

Квадратичная функция потерь

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

ML – решение

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

Градиент

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

Решение

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

где

$$\mathbf{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x_1}) & \dots & \phi_M(\mathbf{x_1}) \\ \phi_0(\mathbf{x_2}) & \dots & \phi_M(\mathbf{x_2}) \\ \dots & \dots & \dots \\ \phi_0(\mathbf{x_N}) & \dots & \phi_M(\mathbf{x_N}) \end{pmatrix}$$

Регуляризация

Функция потерь

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

где (как и раньше)

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

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

Логистическая регрессия

Обобщенные линейные модели

Вопросы