

Линейные методы классификации

Виктор Владимирович Китов
v.v.kitov@yandex.ru

МГУ им.Ломоносова, ф-т ВМиК, кафедра ММП.

I семестр 2015 г.

Регуляризация

- Удобный прием для контроля сложности модели:

$$Q^{regularized}(\mathbf{w}) = Q(\mathbf{w}) + \tau \|\mathbf{w}\|_2$$

$$Q^{regularized}(\mathbf{w}) = Q(\mathbf{w}) + \tau \|\mathbf{w}\|_1$$

$$\|\mathbf{w}\|_1 = \sum_{d=1}^D |w^d|, \quad \|\mathbf{w}\|_2 = \sqrt{\sum_{d=1}^D (w^d)^2}$$

- Регрессия, оцениваемая методом наименьших квадратов с регуляризацией:
 - $\tau \|\mathbf{w}\|_1$ - LASSO
 - $\tau \|\mathbf{w}\|_2$ - Ridge
 - $\alpha \|\mathbf{w}\|_1 + \beta \|\mathbf{w}\|_2$ - elastic net:

Оценка методом максимального правдоподобия

- $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_n\}$ - обучающая выборка,
 $(x_i, y_i) \sim p(y|x, w)$
- Оценка максимального правдоподобия $\hat{w} = \arg \max_w p(Y|X, w)$
- Используя предположение о независимости $y_i|x_i$ $i = 1, 2, \dots, N$:

$$\prod_{i=1}^N p(y_i|x_i, w) = \sum_{i=1}^N \ln p(y_i|x_i, w) \rightarrow \max_w$$

- Оптимизация цены:

$$\sum_{i=1}^n \mathcal{L}(g(x_i)y_i|w) \rightarrow \min_w$$

- Взаимосвязь двух методов:

$$\mathcal{L}(g(x_i)y_i|w) = -\ln p(y_i|x_i, w)$$

Максимальная апостериорная вероятность

- $X = \{x_1, x_2, \dots, x_N\}$, $Y = \{y_1, y_2, \dots, y_N\}$ - обучающая выборка из н.о.р.
($x_i, y_i \sim p(x, y|w)$)
- w - не константа, а с.в. $w \sim p(w)$
- Поскольку

$$p(w|X, Y) = \frac{p(X, Y, w)}{p(X, Y)} = \frac{p(X, Y|w)p(w)}{p(X, Y)} \propto p(X, Y|w)p(w)$$

то справедлива оценка *максимизации апостериорной вероятности*:

$$w = \arg \max_w p(w|X, Y) = \arg \max_w p(X, Y|w)p(w)$$

$$\sum_{i=1}^n \ln p(x_i, y_i|\theta) + \ln p(w) \rightarrow \max_w$$

$\ln p(w)$ соответствует регуляризации.

Варианты априорных вероятностей

- Нормальное распределение

$$\ln p(\mathbf{w}, \sigma^2) = \ln \left(\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\|\mathbf{w}\|_2^2}{2\sigma^2}} \right) = -\frac{1}{2\sigma^2} \|\mathbf{w}\|_2^2 + \text{const}$$

- Распределение Лапласа

$$\ln p(\mathbf{w}, C) = \ln \left(\frac{1}{(2C)^n} e^{-\frac{\|\mathbf{w}\|_1}{C}} \right) = -\frac{1}{C} \|\mathbf{w}\|_1 + \text{const}$$

L_1 норма

- $\|w\|_1$ регуляризация производит отбор признаков.
- Рассмотрим

$$Q(w) = \sum_{i=1}^N \mathcal{L}_i(w) + \lambda \sum_{d=1}^D |w_d|$$

- При $\lambda > \sup_w \left| \frac{\partial \mathcal{L}(w)}{\partial w_i} \right|$ становится заведомо лучше положить $w_i = 0$
- Для более высоких λ больше коэффициентов зануляются.

Логистическая регрессия

- Пусть γ_1, γ_2 - цены неправильной классификации классов ω_1 и ω_2 .
- Предположим

$$\ln \left(\frac{\gamma_1 p(\omega_1 | \mathbf{x})}{\gamma_2 p(\omega_2 | \mathbf{x})} \right) = \beta_0 + \beta^T \mathbf{x}$$

- это эквивалентно

$$p(\omega_2 | \mathbf{x}) = \frac{1}{1 + \exp(\beta'_0 + \beta^T \mathbf{x})}$$

$$p(\omega_1 | \mathbf{x}) = \frac{\exp(\beta'_0 + \beta^T \mathbf{x})}{1 + \exp(\beta'_0 + \beta^T \mathbf{x})}$$

- где $\beta'_0 = \beta_0 - \ln(\gamma_1/\gamma_2)$

Логистическая регрессия

Решающее правило (следуя Байесовскому правилу минимальной цены):

$$x = \begin{cases} \omega_1, & \beta'_0 + \beta^T \mathbf{x} > 0 \\ \omega_2, & \beta'_0 + \beta^T \mathbf{x} < 0 \end{cases}$$

Оценка β'_0, β методом максимального правдоподобия:

$$\prod_{i=1}^N p(c_i | x_i) \rightarrow \max_{\beta'_0, \beta}$$

где c_i - класс объекта x_i .

Многоклассовая логистическая регрессия

- Предположение:

$$\ln \left(\frac{\gamma_s p(\omega_s | \mathbf{x})}{\gamma_C p(\omega_C | \mathbf{x})} \right) = \beta_{s0} + \beta_s^T \mathbf{x}, \quad s = 1, 2, \dots, C-1$$

- Вероятности классов (дающие эквивалентное определение):

$$p(\omega_s | \mathbf{x}) = \frac{\exp(\beta'_{s0} + \beta_s^T \mathbf{x})}{1 + \sum_{s=1}^{C-1} \exp(\beta'_{s0} + \beta_s^T \mathbf{x})}, \quad s = 1, 2, \dots, C-1$$

$$p(\omega_C | \mathbf{x}) = \frac{1}{1 + \sum_{s=1}^{C-1} \exp(\beta'_{s0} + \beta_s^T \mathbf{x})}$$

$$\beta'_{s0} = \beta_{s0} - \ln(\gamma_s / \gamma_C)$$

- Интерпретация: soft-max от дискриминатных функций (для классов $\omega_1, \omega_2, \dots, \omega_{C-1}$) и константы (для класса ω_C).

Многоклассовая логистическая регрессия

- Решающее правило (Байесовское правило минимальной ожидаемой цены):
- $c = \arg \max_c \beta_{c0} + \beta_c^T x$, если $\beta_{c0} + \beta_c^T x > 0$ иначе сопоставить x классу C .
- Оценивание методом максимального правдоподобия:

$$\prod_{i=1}^N p(c_i | x_i) \rightarrow \max_{\beta'_0, \beta}$$

Функция цены

Для 2-х классов $p(y|x) = \sigma(\langle w, x \rangle y)$, где $\sigma = \frac{1}{1+e^{-z}}$,
 $w = [\beta'_0, \beta]$, $x = [1, x_1, x_2, \dots, x_D]$.

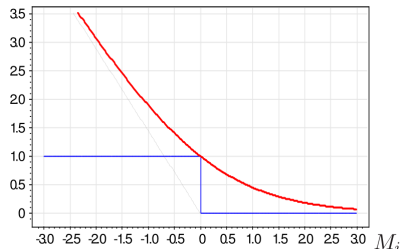
Оценка методом
 максимального правдоподобия:

$$\prod_{i=1}^N \sigma(\langle w, x_i \rangle y_i) \rightarrow \max_w$$

эквивалентна

$$\sum_{i=1}^N \ln(1 + e^{-\langle w, x_i \rangle y_i}) \rightarrow \min_w$$

Следовательно, мажорирующая ф-ция для логистической регрессии $\mathcal{L}(M) = \ln(1 + e^{-M})$.



Метод стохастического градиента

Подставляя $\mathcal{L}(M) = \ln(1 + e^{-M})$ в метод стохастического градиентного спуска, получаем:

$$w \leftarrow w + \eta \sigma(-M_i) x_i y_i$$

Правило обновления перцептрона Розенблатта:

$$w \leftarrow w + \eta \mathbb{I}[M_i < 0] x_i y_i$$

- Обновление логистической регрессии - сглаженный вариант обновления перцептрона Розенблатта.
- Чем существеннее ошибка - тем сильнее обновление весов.

