Машинное обучениее. Линейная классификация.

Московский физико-технический институт, МФТИ

Москва

План лекции

1. Линейная модель классификации;
2. Логистическая регрессия;
3. Метод опорных векторов;
4. Метрики качества классификации;

Основные понятия и обозначения

Дано: выборка обучающих пар обьектов $X^l = (x_i, y_i)_{i=1}^l$.
В общем виде алгоритм классификации представим функцией $a(x, w) = \text{sign}\, f(x, w)$. *Задача*: найти разделяющую поверхность $f(x, w) = 0$. **Отсутпом обьекта** называется величина $M_i(w) = y_i f(x_i, w)$ относительно алгоритма классификации $a(x, w)$.

Аппроксимация эмпирического риска

Пусть - монотонно невозрастающая функция отсупа, мажорирующую функцию потерь $[M < 0] \leq \mathcal{L}(M)$.

$$Q(w, X^l) = \sum_{i=1}^{l} [M_i(w) < 0] \leq \tilde{Q}(w, X^l) = \sum_{i=1}^{l} (M_i(w)) \to \min_{w}.$$

$Q(M) = (1 - M)^2$ - квадратичная,

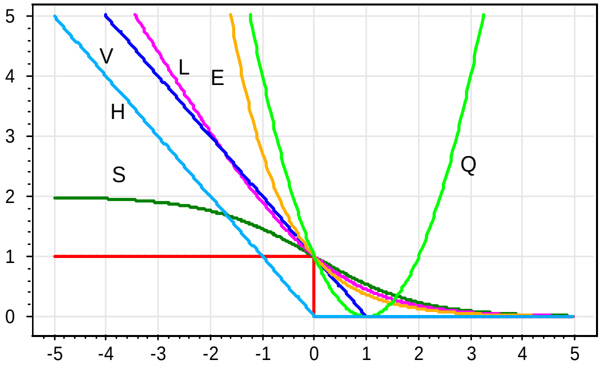
$V(M) = \max\{0, 1 - M\}$ - кусочно-линейная,

$S(M) = 2/(1 + e^{-M})$ - сигмоидная,

$L(M) = \ln_2(1 + e^{-M})$ - логистическая,

$E(M) = e^{-M}$ - экспоненциальная.

Аппроксимация пороговой функции



Мат модель линейной классификации

Рассмотрим классифицирующие модели вида $a(x, w) = \text{sign}\, f(x, w)$, так что множество значений функционала $Y = \{-1, +1\}$. Функция доли неправильных ответов

$$Q(a, x) = \frac{1}{l} \sum_{i=1}^{l} [a(x_i) \neq y_i] = \frac{1}{l} \sum_{i=1}^{l} [\text{sign}\langle w, x_i \rangle \neq y_i] \to \min_{w},$$

или в более компактной записи

$$Q(a, x) = \frac{1}{l} \sum_{i=1}^{l} [y_i \langle w, x_i \rangle <]$$

положим $M_i = y_i \langle w, x_i \rangle$.

Вероятностная модель данных

Применим метод *максимума правдоподобия*. Пусть все наблюдения **независимы** , каждое из которых описывается функцией распределения $p(x, y|w)$, тогда правдопбие выборки можно представить $p(X^l|w) = \prod_{i=1}^{l} p(x_i, y_i|w) \to w$, так что указанный метод эквивалентен постановке минимизации ошибок или функции потерь

$$-\sum_{i=1}^{l} \ln p(x_i, y_i|w)\mathcal{L}(y_i f(x_i, w)).$$

Базовые предоположения

Предположение

Множество прецедентов $X \times Y$ является вероятностным пространством. Выборка прецедентов $X^l = (x_i, y_i)_{i=1}^l$ получена случайно и независимо согласно вероятностному распределению с плотностью $p(x, y) = P_y p_y(x) = P(y|x)p(x)$, где P_y- априорные вероятности, p_y - функции правдоподобия, $P(y|x)$ - апостериорные вероятности классов $y \in Y$.

Предположение

Функции правдопдобия классов принадлежат экспоненциальному семейству плотностей, имеют равные значения параметров d и δ, но отличаются значениями параметров сдвига θ_y.

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN},$$

$$precision = \frac{TP}{TP + FP},$$

$$recall = \frac{TP}{TP + FN},$$

Критерий качества на основе точности и полноты: F - мера, гармоническое среднее точности и полноты:

$$F = \frac{2 * precision * recall}{precision + recall}$$

Площадь под ROC -кривой

Доли неверно принятых обьектов (False Positive Rate) и верно принятых обьектов (True Positive Rate)

$$FPR = \frac{FP}{FP + TN},$$

$$FPR = \frac{TP}{TP + FN}.$$

Проблема чувствительности к соотношению классов -> решение через precision-recall кривая оценки.

Пример ROC -кривой