Home Work 4. Theory.

Зверева Дарья

12 апреля 2017 г.

3 Теоретическое задание

3.1 Знакомство с линейным классификатором

1 Как выглядит линейный классификатор

$$a(x) = \begin{cases} 1, & \text{если } f(x) > 0 \\ -1, & \text{если } f(x) < 0 \end{cases}$$

$f(x) = \omega_0 + \omega_1 x_1 + \ldots + \omega_n x_n = \langle \omega, x \rangle + \omega_0$

2 Что такое отступ алгоритма на объекте? Отступом называется величина $M_i = y_i f(x_i)$, где y_i – класс, к которому относится объект x_i.

Если $M_i > 0$, то классификатор $a(x)$ дал верный ответ, если $M_i \leqslant 0$, то классификатор ошибся.

4 Что такое функционал эмпирического риска? Назовём функционалом эмпирического риска $Q(\omega, X^l) = \sum\limits_{i=1}^{l} [M_i(\omega) < 0]$.

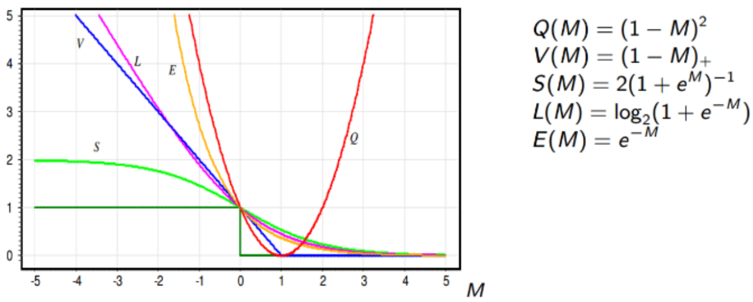
Для наилучшего классификатора он должен принимать минимальное значение.

6 Что такое функционал аппроксимированного эмпирического риска? Введем функцию потерь $\mathcal{L}(M_i(w))$ – монотонно невозрастающая функция, так чтобы $[M < 0] \leqslant \mathcal{L}(M)$.

Назовём функционалом эмпирического риска $Q(\omega, X^l) = \sum\limits_{i=1}^{l} \mathcal{L}(M_i(\omega))$.

Для наилучшего классификатора он должен принимать минимальное значение.

7 Что такое функция потерь? Функция потерь - неотрицательная функция $\mathcal{L}(M)$, характеризующая величину ошибки алгоритма a на входе x. На рисунке представлены графики самых распространённых функций потерь.



8 Пример негладкой функции потерь. Как пример можно видеть тёмно-зелёную и синюю функции с графика, расположенного выше.

$$F(M) = \begin{cases} 1, & \text{если } M > 0 \\ -1, & \text{если } M < 0 \\ [-1; 1], & \text{если } M = 0 \end{cases}$$

или $V(M) = (1 - M)_+$

9 Что такое регуляризация? Некоторое добавляемое ограничение на коэффициенты. Оно нужно для того, чтобы модель не переобучалась.

$Q(\omega, X^l) = \sum\limits_{i=1}^{l} \mathcal{L}(M_i(\omega)) \to \min$

l_1 регуляризация: $\sum\limits_{k=1}^{m} |w_k| \leqslant \tau$

l_2 регуляризация: $\sum\limits_{k=1}^{m} w_k^2 \leqslant \tau$

15 Что представляют собой метрики качества Accuracy, Precision, Recall? tn - количество истинно-негативных срабатываний tp - количество истинно-положительных срабатываний fp - количество ложно-положительных срабатываний fn - количество ложно-негативных срабатываний

$Accuracy = \frac{tp+tn}{tp+tn+fp+fn}$ - доля правильных ответов при классификации.

$Precision = \frac{tp}{tp+fp}$ - точность

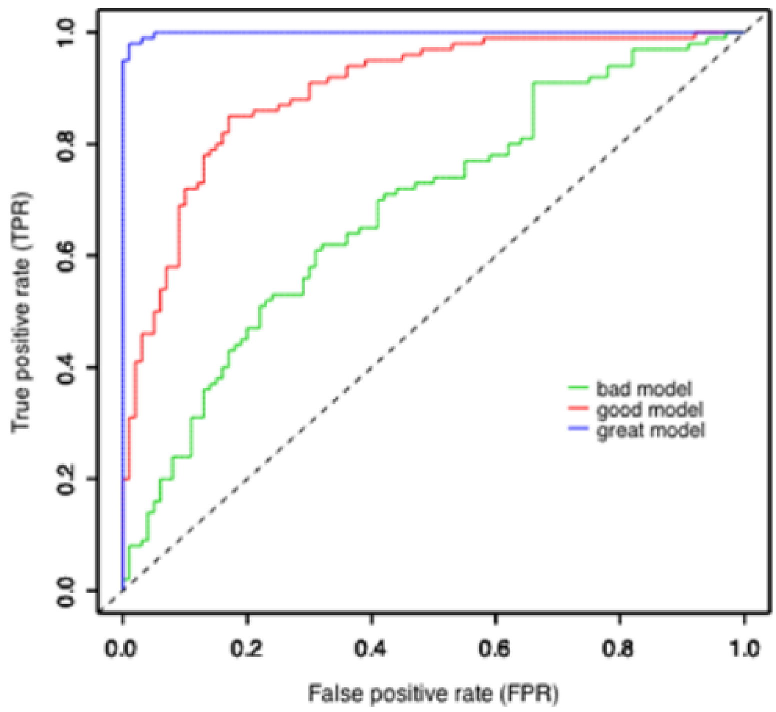
$Recall = \frac{tp}{tp+fn}$ - полнота

16 Что такое метрика качества AUC и ROC-кривая? Если мы введём следующие обозначения:

$TPR = \frac{tp}{tp+fn}$

$FPR = \frac{fp}{fp+tn}$

И построим график TPR(FPR)



То метрика качества AUC = area under curve - площадь под ROC-кривой.

17 Как построить ROC-кривую?

1. вычислить количество представителей классов +1 и -1 в выборке: $m_- := \sum\limits_{i=1}^{m}[y_i = -1]$, $m_+ := \sum\limits_{i=1}^{m}[y_i = +1]$
2. упорядочить выборку X^m по убыванию значений $f(x_i, w)$;
3. установить начальную точку ROC-кривой: $(FPR_0, TPR_0) := (0, 0)$; $AUC := 0$;
4. $\forall i = 1..m$ если $(y_i = -1)$, то сместиться на один шаг вправо: $FPR_i := FPR_{i-1} + \frac{1}{m_-}$; $TPR_i := TPR_{i-1}$; $AUC+ = \frac{1}{m_-}TPR_i$;
5. иначе сместиться на один шаг вверх: $FPR_i := FPR_{i-1}$; $TPR_i := TPR_{i-1} + \frac{1}{m_+}$;

3.3 SVM и максимизация разделяющей полосы

Есть линейный классификатор $a(x) = sign(\langle\omega, x\rangle - \omega_0)$

Использует кусочно-линейную функцию потерь и l_2-регуляризатор:
$\sum\limits_{i=1}^{l} \mathcal{L}(M_i) + \gamma ||\omega||^2 \to \min\limits_{\omega}$

$\mathcal{L}(M_i) = \max 0, 1 - M_i = (1 - M_i)_+$

Хотим построить разделяющую плоскость и максимизировать зазор:

$$\begin{cases} \langle \omega, \omega \rangle \to \min; \\ y_i \left(\langle \omega, x_i \rangle - \omega_0 \right) \geqslant 1, i = 1, \dots, l \end{cases}$$

В случае линейно неразделимых классов:

$$\begin{cases} \frac{1}{2} \langle \omega, \omega \rangle + C \sum\limits_{i=1}^{l} \xi_i \to \min\limits_{\omega, \omega_0, \xi}; \\ y_i \left(\langle \omega, x_i \rangle - \omega_0 \right) \geqslant 1 - \xi_i, i = 1, \dots, l \\ \xi_i \geqslant 0, i = 1, \dots, l \end{cases}$$

3.4 Kernel Trick

Попробуем ядро $K(x, y) = \langle x, y \rangle^2 = (x_1 y_1 + x_2 y_2)^2 = x_1^2 y_1^2 + 2 x_1 x_2 y_1 y_2 + x_2^2 y_2^2 = \langle (x_1^2, x_2^2, \sqrt{2} x_1 x_2, y_1^2, y_2^2, \sqrt{2} y_1 y_2) = \langle \Phi(x), \Phi(y) \rangle$.

$\Phi(x) = (x_1^2, x_2^2, \sqrt{2} x_1 x_2)$

Подберём веса:

$\omega_0 + \Phi(x) \omega = \omega_0 + \omega_1 x_1^2 + \omega_2 x_2^2 + \omega_3 \sqrt{2} x_1 x_2$.

$\omega_0 = -3, \omega = (1, 2, 0)$.

3.6 Метрики качества

Что представляют собой метрики качества Accuracy, Precision, Recall?

tn - количество истинно-негативных срабатываний tp - количество истинно-положительных срабатываний fp - количество ложно-положительных срабатываний fn - количество ложно-негативных срабатываний

$Accuracy = \frac{tp+tn}{tp+tn+fp+fn}$ - доля правильных ответов при классификации.

$Precision = \frac{tp}{tp+fp}$ - точность

$Recall = \frac{tp}{tp+fn}$ - полнота

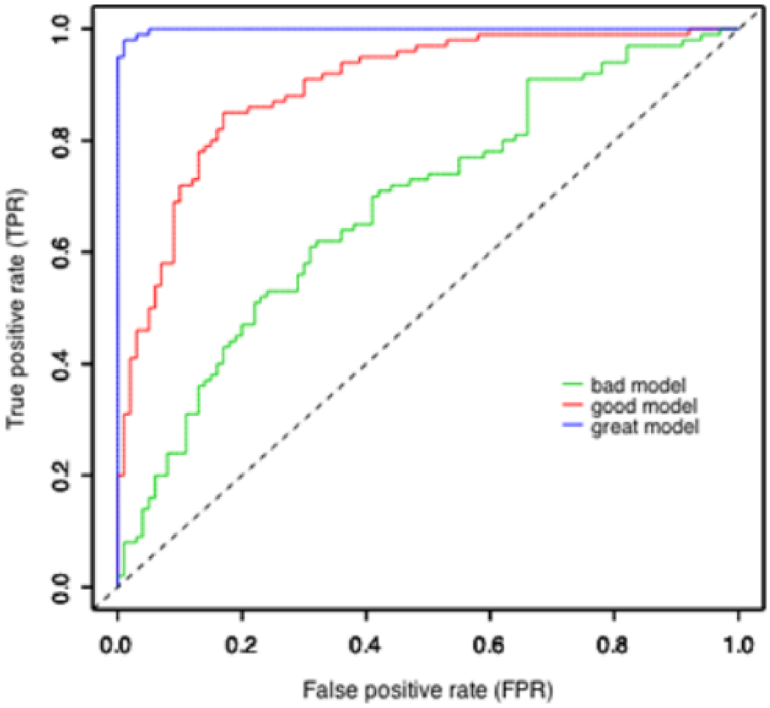
Что такое метрика качества AUC и ROC-кривая?

Если мы введём следующие обозначения:

$TPR = \frac{tp}{tp+fn}$

$FPR = \frac{fp}{fp+tn}$

И построим график TPR(FPR)



То метрика качества AUC = area under curve - площадь под ROC-кривой.

Как построить ROC-кривую?

1. вычислить количество представителей классов +1 и -1 в выборке: $m_- := \sum\limits_{i=1}^{m}[y_i = -1]$, $m_+ := \sum\limits_{i=1}^{m}[y_i = +1]$
2. упорядочить выборку X^m по убыванию значений $f(x_i, w)$;
3. установить начальную точку ROC-кривой: $(FPR_0, TPR_0) := (0, 0)$; $AUC := 0$;
4. $\forall i = 1..m$ если $(y_i = -1)$, то сместиться на один шаг вправо: $FPR_i := FPR_{i-1} + \frac{1}{m_-}$; $TPR_i := TPR_{i-1}$; $AUC + = \frac{1}{m_-} TPR_i$;
5. иначе сместиться на один шаг вверх: $FPR_i := FPR_{i-1}$; $TPR_i := TPR_{i-1} + \frac{1}{m_+}$;