Теоретические задачи

3.1 Линейный классификатор

1. $\mathbb{X} = \mathbb{R}^d$ - пр-во объектов

$\mathbb{Y} = \{-1; 1\}$ - мн-во допустимых ответов

$X = \{(x_i, y_i)\}_{i=1}^{\ell}$ - обучающая выборка

Лин. модель: $a(x) = \text{sign}(\overbrace{\langle w, x\rangle + w_0}^{\text{дискриминантная}}) = \text{sign}\left(\sum_{j=1}^{d} w_j x_j + w_0\right)$

Геом. смысл: гиперплоскость с вектором нормали $w$.

$\langle w, x\rangle$ - расстояние от гиперплос-ти до $x$
знак - с какой стороны от гиперплоскости находится данная точка

2. Введем функционал ошибки:

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

Функционал дискретен отн весов $\Rightarrow$ минимум искать с помощью градиентного спуска невозможно.
Также может быть много минимумов

$$\Rightarrow Q(a, X) = \frac{1}{\ell}\sum_{i=1}^{\ell}[\underbrace{y_i\langle w, x_i\rangle}_{M_i} < 0] \to \min$$

$M_i = y_i\langle w, x_i\rangle$ - отступ

знак: корректность ответа классификатора
абс. величина: степень увер-ти классификатора в ответе

③ Введении константного признака $x_0 = 1$, а к вектору весов $w_0$.

④ $\tilde{Q}(w) = \sum_{i=1}^{l} [M_i(w) < 0] \to \min_{w}$ – функционал эмпирического риска, $M_i = [a(x_i) \neq y_i)$

Для наилучшего алгоритма класс принимает 0.

⑤ $\tilde{Q}(w) = \sum_{i=1}^{l} [M_i < 0] \to \min_{w}$

при $w = 0$.

⑥ $Q(a, X) = \frac{1}{l} \sum_{i=1}^{l} L(M)$

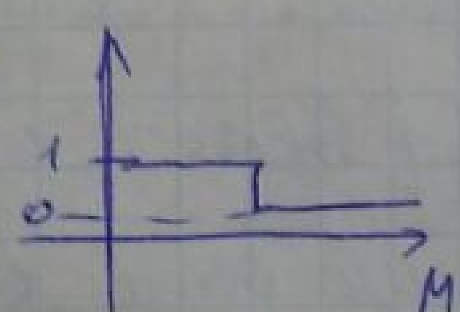
⑦ Ф-ия потерь – способ измерения ошибки алгоритма

Для лин классиф.

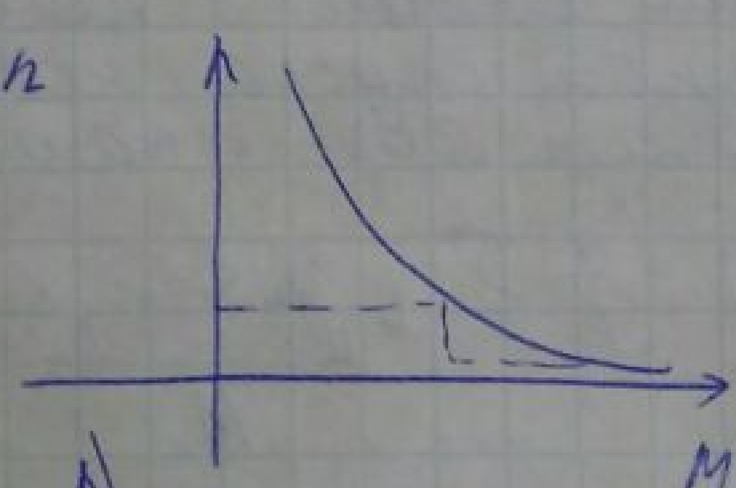
$Q(a, X) = \frac{1}{l} \sum_{i=1}^{l} [a(x_i) \neq y_i]$.

Обычно выпуклая в случ. линейной классификации невозрастающая

Пороговая ф.п




⑧ Экспоненц. ф.п




Логистическ

⑧ Нежадкая ф-ия потерь — пороговая ф-ия потерь $L(M) = [M \le 0]$

⑨ Регуляризация — способ борьбы с переобучением, штраф за сложность
Usage: добавление некоторой ф-ии в минимизирующую ф-ию.

$$Q(w, X) + \underbrace{\lambda}_{\text{коэффициент регуляризации}} \|w\|^2 \to \min$$

- $L_1$-регуляризатор : $\|w\|_1 = \sum |w_j|$
- $L_2$-регуляризатор: $\|w\|_2 = \sum w_j^2$

⑩ Переобучение — алгоритм подогнан под объекты обуч-ей выборки

При попытке обобщения алгоритма на новые данные, показ-ся плохой результат.

Регуляризация — способ ~~его~~ борьбы с переобучением ⇒

повышение обобщающей способности алгоритма

⑪ Эмпирический риск — средняя величина ошибки алгоритма на обучающей выборке

$$Q(a, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} \mathcal{L}(a, x_i) \text{ — функционал эмпирического риска}$$

т.к. вычисл-ся по эмпирическим данным

Метод, минимизирующий эмпирический риск, может получить обучающую выборку, запомнить её и строит алгоритм, который сравнивает предъявляемый объект с обучающими. Такой алгоритм не способен восстановить зависимость вне материала обучения. ⇒ переобучение

При отступе от точки минимума, значение функционала аппроксимированного эмпирического риска сильно возрастает

(12) Если параметры выходят из допустимых значений => регуляризация увеличивает зн-ие аппр. риск.

(13) Регуляризация влияет на значение функционала (увеличивает) => с регуляризацией.

(14) С одной стороны, без регуляризации, т.к без неё происходит переобучение, и след-но, плохие результаты

С другой стороны, с регуляризацией, при случае, когда даже значение с переобучением было меньше, чем если бы мы ~~добавили~~ добавили регуляризатор

(15) Есть бинарная классификация:

| | $y = 1$ | $y = -1$ |
|---|---|---|
| $1 = a(x)$ | True Positive (TP) | False Positive (FP) |
| $-1 = a(x)$ | False Negative (FN) | True Negative (TN) |

— ответ реальный

ответ алгоритма

TP, FN, FP, TN – кол-во объектов

$$\text{Accuracy} = \frac{TP + TN}{X_1 + X_{-1}}$$, где $X_1$ – кол-во объектов класса $\{1\}$, аналогично $X_{-1}$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{P}$$

(16) Пусть $FPR = \frac{FP}{X_{-1}}$ – False Positive Rate

$TPR = \frac{TP}{X_1}$ – True и – "

ROC-кривая – кривая точности-полноты т.е. зав-ть TPR(FPR) – доли верных положит. классифик. от доли ложных положительных

AUC-метрика – площадь под ROC-кривой

(17) Дано: выборка

ф-ия: $f(x, w)$ – дискриминантная

Output: $\{(FPR_i, TPR_i)\}_{i=0}^{m}$ – пос-ть точек ROC-кривой

Алгоритм:

1) вычислить представит. классов:

$m_-$ и $m_+$

2) Упорядочим выборку по убыванию зн-ий $f(x_i, w)$

3) $(0, 0)$ – изначальная точка ROC-кривой

4) $\forall i = 1..m$

• если $m_-$, то сместиться на 1 шаг вправо

~~$FPR_i = FPR_{i-1} + \frac{1}{m_-}$~~

$$(TPR_i, FPR_i) = (TPR_{i-1}, FPR_{i-1} + \frac{1}{m_-})$$

• иначе сместиться вверх

$$(TPR_i, FPR_i) = (TPR_{i-1} + \frac{1}{m_+}; FPR_i)$$

## №3.2. Вероятностный смысл регуляризаторов

Введем вероятностное пространство и совместную плотность распределения объектов и классов $p(x, y | w)$

$\left.\begin{matrix} \gamma\text{-гиперпараметр} \\ w \end{matrix}\right\}$ - параметры модели

введем априорную плотность $p_\gamma(w)$.

Запишем ф-ию правдоподобия выборки X и модели

$$F_\gamma(w, X) = \ln \cdot p_\gamma(w, X) =$$

$$= \sum_{i=1}^{\ell} \ln p(x_i, y_i | w) + \ln p_\gamma(w) \to \max$$

Миним. ф-ия потерь аппроксимированного эмпирического риска экв.-на ф-ией максимиз-ии максимальн правдоподобия:

$$Q(w, X) = \sum_{i=1}^{\ell} \ln p(x_i, y_i | w) + \ln p_\gamma(w) =$$

$$\left| \begin{matrix} -\sum \ln p(x_i, y_i | w) = \sum L(y, f) \\ -\ln p_\gamma(w) = \gamma \cdot V - \text{регуляризатор} \end{matrix} \right| =$$

$$= \sum L(y, f) + \gamma \cdot V$$

$\Rightarrow$ регуляриз. имеет апр. распр-ие некоторого семейства плотностей.

$\ell_1$: распр-ие Лапласа

$\ell_2$: n-мерное гауссовское

## №3.4 Kernel trick

Рассмотрим следующее ядро:

$K(x, y) = \langle x, y \rangle^2 = x_1^2 y_1^2 + 2x_1 y_1 x_2 y_2 + x_2^2 y_2^2 =$

$= \langle (x_1^2, x_2^2, \sqrt{2} x_1 x_2, y_1^2, y_2^2, \sqrt{2} y_1 y_2) \rangle$

$\Rightarrow$ ~~...~~ размерность $\mathbb{R}^3 \Rightarrow$

отобр-ие $(x_1, x_2) \to (x_1^2, x_2^2, \sqrt{2} x_1 x_2)$

Из условие разделяющая поверхность имеет координаты: $w = (1, 2, 0)$, $w_0 = -3$

$(x_1, x_2, \sqrt{2} x_1 x_2) \cdot (1, 2, 0) + w_0 = x_1^2 + 2x_2^2 - 3 = 0$

Как в условии

## №3.3. SVM и максимизация разделяющей полосы

$a(x) = sign(\langle w, x \rangle - w_0)$