Задание 1

Пусть дана выборка X и классификатор $b(x)$, предсказывающий оценку принадлежности объекта к положительному классу. Постройте roc-кривую и вычислите AUC-ROC для множества классификаторов $a(x,t)$ порожденных $b(x)$ на выборке X

	$b(x)$	y
x_8	0.95	1
x_5	0.9	1
x_2	0.8	1
x_7	0.6	0
x_6	0.3	1
x_4	0.25	0
x_3	0.2	0
x_1	0.1	1

Порог 1: $t=0{,}96$ $FPR=0$ $TPR=0$

Порог 2: $t=0{,}92$ $FPR=0$ $TPR=0.2$

Порог 3: $t=0{,}85$ $FPR=0$ $TPR=0{,}4$

Порог 4: $t=0{,}7$ $FPR=0$ $TPR=0.6$

Порог 5: $t=0.5$ $FPR=\frac{1}{3}$ $TPR=0.6$

Порог 6: $t=0{,}28$ $FPR=\frac{1}{3}$ $TPR=0{,}8$

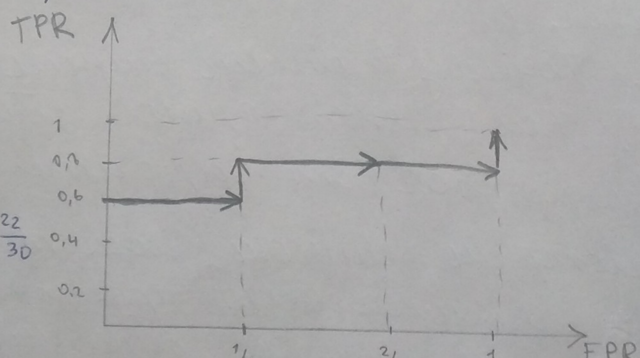
Порог 7 $t=0{,}22$ $FPR=\frac{2}{3}$ $TPR=0{,}8$

Порог 8 $t=0{,}15$ $FPR=1$ $TPR=0{,}8$

Порог 9: $t=0{,}05$ $FPR=1$ $TPR=1$

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

$$TPR=\frac{FP}{TP+FN}$$



$$AUC\text{-}ROC=\frac{1}{3}\cdot 0{,}6+\frac{2}{3}\cdot 0{,}8=0{,}2+\frac{8}{15}=\frac{22}{30}$$

Задача 3 Пусть дана некоторая выборка X и классификатор $b(x)$, возвращающий в качестве оценки принадлежности объекта к положительному классу 0 или 1 (а не некоторое вещественное число)

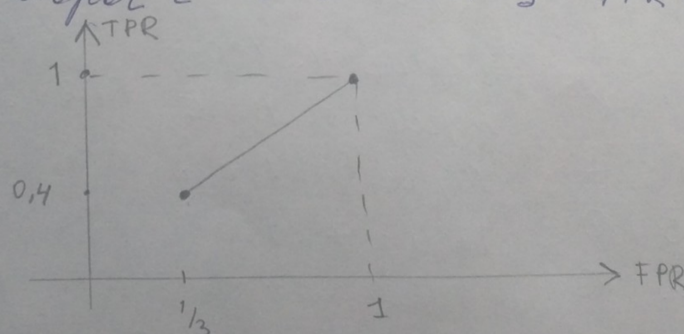
1. Построить ROC-кривую для классификатора $b(x)$ на выборке X

Для примера возьмем выборку из задания 1:

	$b(x)$	y
x_1	1	1
x_2	1	1
x_3	1	1
x_4	1	0
x_5	0	1
x_6	0	0
x_7	0	0
x_8	0	1

Порог 1: $t=1$ $FPR = 1/3$ $TPR = 0{,}4$

Порог 2: $t=0$ $FPR = 1$ $TPR = 1$



Задание 8

Задача 3.

2. Почему логистическая регрессия позволяет предсказывать корректные вероятности принадлежности объекта классам?

Метод максимального правдоподобия позволяет корректно восстанавливать функции распределения.

3. Рассмотрим оптимизационную задачу hard-margin SVM. Всегда ли в обучающей выборке существует объект x_i, для которого выполнено условие $y_i(\langle w, x_i \rangle + b) = 1$?

Из условия $\begin{cases} \frac{1}{2}\|w\|^2 \to \min\limits_{w,b} \\ y_i(\langle w, x_i \rangle + b) \geq 1, \ i = 1 \ldots \ell \end{cases}$

следует, что условие $y_i(\langle w, x_i \rangle + b) = 1$ не обязано выполняться

Задача 3.

4. С какой целью в постановке оптимизационной задачи soft-margin SVM вводятся переменные $\xi_i, i=\overline{1,l}$

Линейно неразделимый случай: условия неравенства из hard-margin SVM не будут выполнены.

Поэтому сформулируем задачу таким образом, чтобы за попадание внутрь разделяющей полосы объекты получают штраф ξ_i:

$$\begin{cases} \frac{1}{2} \|w\|^2 + C^2 \sum_{i=1} \xi_i \to \min\limits_{w,b,\xi}, \\ y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i, \ i=1,\dots,l, \\ \xi_i \geq 0, \ i=1,\dots,l \end{cases}$$

Задача 7. Вычислите градиент функции $L(x, y; w)$ для случая линейного классификатора

$$L(x,y;w) = \log(1+\exp(-y\langle w, x \rangle))$$

В ответе использовать сигмоидную функцию

$$\sigma(z) = \frac{1}{1+\exp(-z)}$$

Проверить, что $\sigma'(z) = \sigma(z)(1-\sigma(z))$

1. Проверим, что $\sigma'(z) = \sigma(z)(1-\sigma(z))$

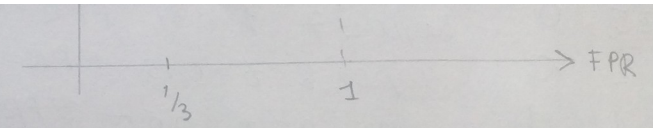
$$\frac{d\sigma}{dz} = \frac{d}{dz}\frac{1}{1+\exp(z)} = -\frac{-\exp(z)}{(1+\exp(-z))^2} = \frac{1}{1+\exp(-z)} \cdot \frac{\exp(-z)+1-1}{1+\exp(-z)} = \sigma(z)(1-\sigma(z))$$

2. $L(x,y,w) = \log(1+\exp(-y\langle w,x \rangle)) = -\log(\sigma(y\langle w,x \rangle))$

$$\frac{\partial L}{\partial w} = -\frac{1}{\sigma(y\langle w,x \rangle)} \cdot \sigma'(y\langle w,x \rangle) = \frac{1}{\sigma(y\langle w,x \rangle)} \cdot \sigma(y\langle w,x \rangle)(1-\sigma(y\langle w,x \rangle))yx$$

$$= yx(1-\sigma(y\langle w,x \rangle))$$

8 0 1



Задание 8.

1. Почему в общем случае распределение $p(y|x)$ для некоторого объекта $x \in \mathbb{X}$ отличается от вырожденного ($p(y|x) \in \{0,1\}$)?

Рассмотрим случай логистической регрессии. Метод максимального правдоподобия будет использован для ее обучения.

Возьмем $y \in Y$ в качестве случайной величины, которая имеет распределение Бернулли с вероятностью p, где $p = z(\langle w, x \rangle)$, $x \in \mathbb{X}$

Т.е. мы восстановим функцию вероятн.

$$p(Y=y \mid X=x) = z(\langle w, x \rangle)^y \left(1 - z(\langle w, x \rangle)\right)^{1-y}$$

Таким образом, классификатор будет обучаться давать те ответы, которые были бы представлены случайными величинами с распр. $Ber(p)$, но на объектах обучающей выборки.

Задание 8.

3. Рассмотрим оптимизационную задачу hard-margin SVM. Всегда ли в обучающей выборке существует объект x_i, для которого выполнено $y_i(\langle w, x \rangle + b = 1$? Почему?

Да, так как это условие соответствует принадлежности точки к полосе разделения. Алгоритм SVM максимизирует зазор между гиперплоскостью и объектами классов, расположенных наиболее близко к ней.
Но все равно в каждом из классов найдётся как минимум один граничный объект обучающей выборки, отступ которого равен этому минимуму: иначе можно было бы сдвинуть гиперплоскость в сторону класса с большим отступом, и тогда расстояние от гиперплоскости до объектов обучающей плоскости увеличится.

4. С какой целью в постановке оптимизационной задачи soft-margin SVM вводятся переменные $\varepsilon_i, i = \overline{1, l}$?

Данные переменные являются штрафом для объектов x, что позволяет решать задачи с линейно неразделимыми множествами.