№1.

Блюменау М.И.
БФЗ191
ИАД-6

1)

$a(x) = \langle w, x \rangle$

Я наивно предполагаю что тут я уже добавил константный признак (для $\forall$ объекта $x_0 = 1$, соотв. вес w_0). $\langle \ \rangle$ — скалярное произв., w — веса модели (их подбираем), x — объект, $a(x)$ выдает ответ

N — признаков

Иной вид записи

$$a(x) = w_0 + \sum_{i=1}^{N} w_i x_i$$

2)

	Признак	Ответ
а)	0	10
б)	2	20
в)	4	30

$w_0 = 10$

$w_1 = 5$

Тогда

$w_1 x + w_0$

а) $0 \cdot 5 + 10 = 10$

б) $2 \cdot 5 + 10 = 20$

в) $4 \cdot 5 + 10 = 30$

Ошибка 0, как ни крути.

Без w_0:

$w_1 x$, в $x = 0$ ответ всегда 0 ($0 \cdot$ любое число $= 0$)

3) 90 признаков дадут 90 параметров

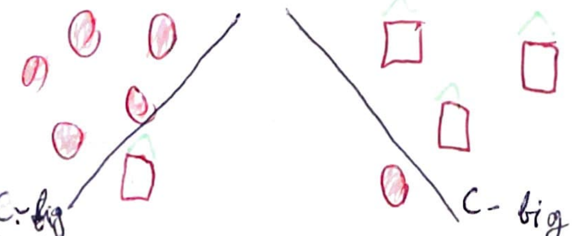
$+ w_0$ (константный) $\rightarrow$ 91

Если добавить L_2, то параметров станет в 2 раза больше (только надо не забыть, что w_0 не регуляриз., поэтому его не учитываем)

т.е.: $90 \cdot 2 + 1 = 181$

№2.

1) Гиперпараметр C влияет на то, насколько нам важно классификация наших групп говоря. Нарисую картинку:



(задача о разделении домиков и колодцев)

Ну мы как бы выбираем между отступам

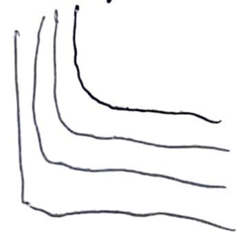
и классиф.. Если можно разделить прямой однозначно, такой проблемы нет.

Ну и из формулки очевидно, что при $C = 0$ ξ_i любые, ибо не влияют на min (1 строчка в $\{$) и мы получим модель, которая будет в зависимости от ξ_i давать/нет послабления на отступы от ближ. объектов 0%,

2) Потому что среднее арифметическое даст невесть что. Грубо говоря, мы получим

Ср. арифм.

F - мера



И как бы из картинок (я не эксперт, warning!!!), видно, что для ср. арифм $rec. = 0{,}4$ и $pr. = 0{,}6$ то же самое, что $rec. = 0{,}2$, $pr. = 0{,}9$, а зачем нам так надо?

F-мера же как раз тоже способствует равенству наших pr. и rec (при их одновременном) увеличении. Т.е. $F = \frac{2\, rec \cdot pr}{pr + rec}$. Есть F_β-мера $= (1+\beta^2)\frac{pr \cdot rec}{\beta^2 rec + pr}$,

$\beta < 1$ - важна точность

$\beta > 1$ - полнота

3) Ну зависит от нашей задачи конечно. Но всегда работает перебор.

Отсортируем скалярные произведения и просто начнём идти, меняя t (можно grid search ещё заюзать)

4) Есть One-vs-all и All-vs-all

K-классов $X_K = (x_i, [y_i = K])_{i=1}^{\ell}$ $\quad Y = \{1, \dots, K\}$

обучаем $a_K(x)$ на X_K

$a(x) = \arg\max\limits_{k=1\dots K} a_k(x)$ (Просто как было на лекции: можно решить задачу для каждой пары классов и разделяться потом с объединением этого)

All-vs-all:

$X_{km} = \{(x_i, y_i) \in X \mid y_i = k$ или $y_i = m\}$ $\quad a(x) = \arg\max \sum\limits_{m=1}^{K} [a_{km}(x) = k]$

1) Жадный алгоритм: № 3.

0) ~~остаток~~ = выход

1) ищем такой предикат, что $j_t = \arg\min Q(R_m, j)$ выполн.

3) Разбиваем так, что

$R_p = \{(x, y) \in R_m \mid [x_j \geq t]\}$ ← тута все у кого значение $\geq t$

$R_e = \{ (x, y) \in R_m \mid [x_j < t]\}$ ← тута все, у кого $< t$

4) Повторяем пока не выйдем

2) В листьях хранятся наши прогнозы (или ответы, кому как нравится)

Их можно взять усреднением, типо $C_0 = \frac{1}{|R_0|} \sum\limits_{(x_i, y_i) \in R_0} y_i$

Можно туда засунуть лин модель

$C_0(x) = \langle w_0, x \rangle$

4)

Переобученная - 1. Тогда мы буквально подгоняемся под наши ответы

Максимальную глубину - если в плане того, что она что-то всё же решает, то число объектов в мин. классе

Если никто не требует слишком работать - хватит все объекты выборки :) (но это недообучение)

3) Потому что наверное количественная дерево, которое "режет" плоскость параллельно осям, может сделать нам "лесенку" и будет очень напоминать лин. рег., типо

5) Энтропия также если у меня равномерное распределение

№4.

w^0

$w^t = 0,5\, w^{t-1} - \frac{1}{t^{0,5}} \nabla(w^{t-1} + w^{t-2})$

Останов.

$\left\| \frac{w^t}{w^{t-1}} \right\| < 0,01$

1) Идея делать нули в w^0 не ошибка, но рандом будет лучше

2) Откуда 0,5 перед w^{t-1}? Мы так и берём и начинаем каждый раз делить на 2 значения w, классно, уйдём в 0.

3) Зачем нам $w^{t-1} + w^{t-2}$? Там нужен w^{t-1} только, ведь мы идём по шагам, а тут сумма двух прошлых ...

4) ~~Останов~~. Останова.

Писал гений, кажется в итоге у нас w^t и w^{t-1} почти одинаковы (если сошлись), а мама говорила $\frac{x}{x} = 1$, поэтому тут не $< 0,01$, а почти $= 1$ (или точно $= 1 - \varepsilon$)

5) $\frac{1}{t^{0,5}}$ может быть не очень. Мы видели несколько раз, что числа и степени и числителя небольшие (на лекции было что-то типа $0,1/t^{0,3}$), но тут это может и сработать

$\boxed{4/}$

№5.

$b(x)$	0	2	1	4	3	10	11	-2
y	-1	-1	+1	+1	+1	-1	+1	-1

AUC-ROC:

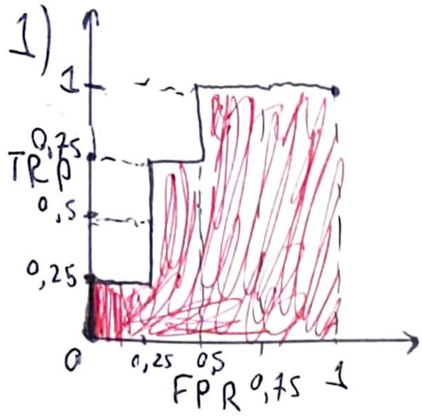
Красным на картинке

Просто сумма прямоугольников:

$0{,}25\cdot 0{,}25 + 0{,}25\cdot 0{,}75 + 0{,}25\cdot 1 + 0{,}25\cdot 1 =$

$= 0{,}0625 + 0{,}5 + 0{,}1875 = 0{,}75$, что вроде норм

1)



ROC-кривая

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

$Y: TPR = \frac{TP}{TP+FN}$

Левая (0, 0)

Правая (1, 1)

2) $t = 5$:

pr. $= \frac{TP}{TP+FP} = \frac{1}{1+1} = 0{,}5$

rec. $= \frac{TP}{TP+FN} = \frac{1}{1+3} = 0{,}25$

Ну наверное надо отсортировать всё еще:

	w	и	з	д	г	в	б	а	
$b(x)$	-2	0	1	2	3	4	10	11	
y	-1	-1	+1	-1	+1	+1	-1	+1	0,0

а) $\frac{0}{\ldots} = 0$

$\frac{1}{1+3} = \frac{1}{4}$

б) $\frac{1}{1+3} = \frac{1}{4}$

$\frac{1}{1+3} = \frac{1}{4}$

в) $\frac{1}{1+3} = \frac{1}{4}$

$\frac{2}{2+2} = 0{,}5$

г) $\frac{1}{2+3} = \frac{1}{4}$

$\frac{3}{3+1} = 0{,}75$

д) $\frac{2}{2+2} = 0{,}5$

$\frac{3}{3+1} = 0{,}75$

з) $\frac{2}{2+2} = 0{,}5$

$\frac{4}{4+0} = 1$

и) $\frac{3}{3+1} = 0{,}75$

$\frac{4}{4} = 1$

w) $\frac{4}{4} = 1$

$\frac{4}{4} = 1$

5/

3) ~~Ну или, только что в п. 2 получили ровно это.~~

Очевидно из сортировки $t < 10{,}5$

4) Смотрим на ROC, видим, что на оси y сидит наша полнота (rec.), смотрим когда 50% первый раз: это наш b). При ней полнота 50%, точность:

$$\frac{2}{2+1} = \frac{2}{3} \approx 0{,}67$$

это 11 и 4 ← этот в 10 сидит

Ну и порог 13,5 нас устроит получается