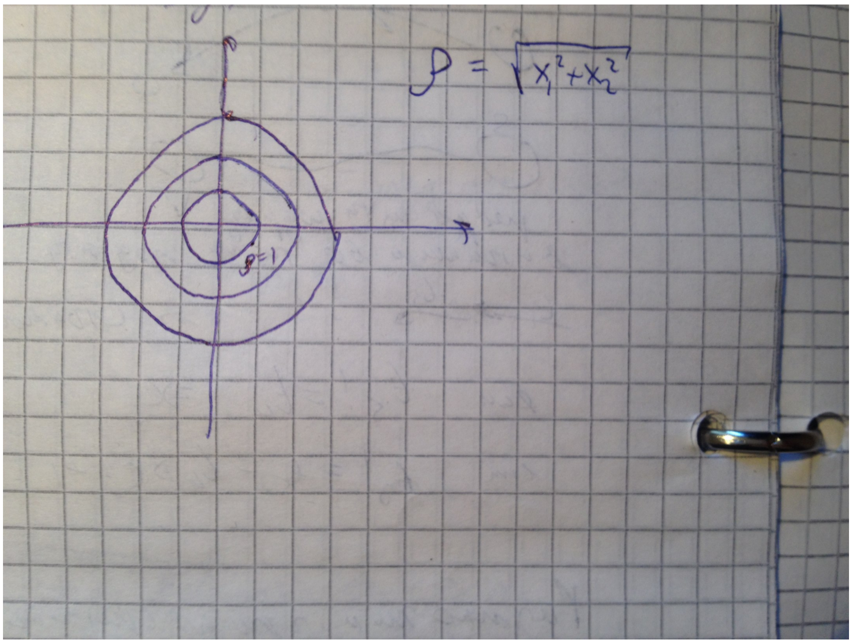
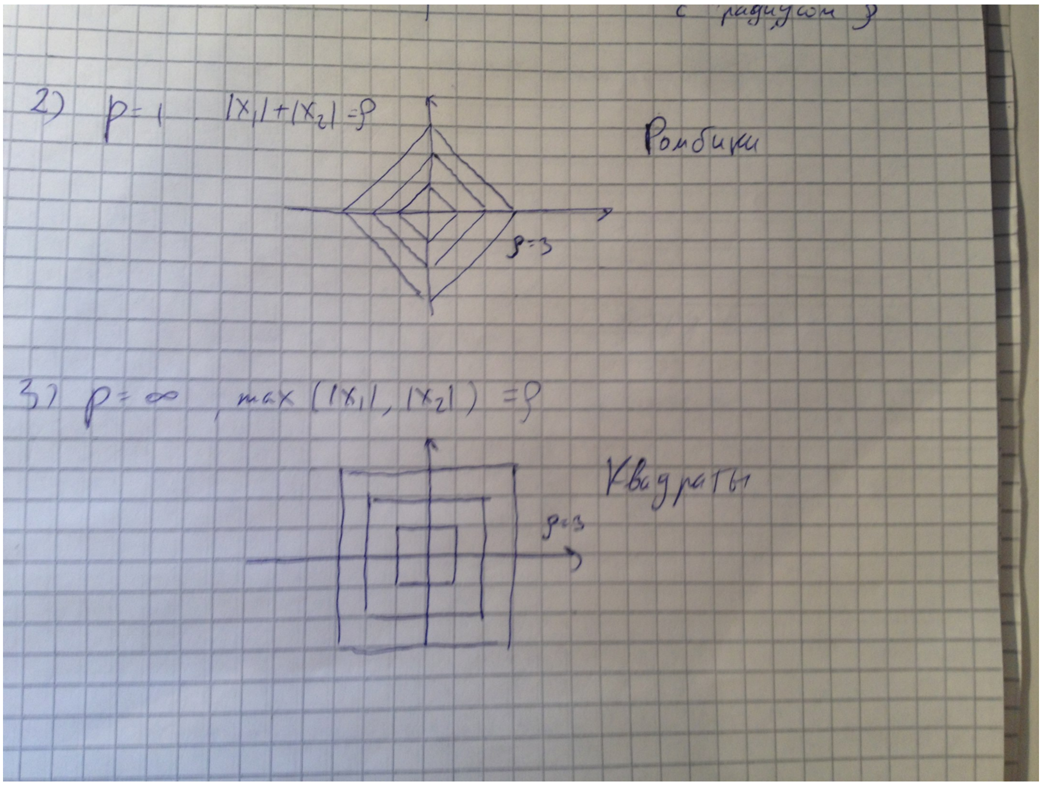
Задание 1.
1) т.к. выборка имеет малый размер, то имеет смысл использовать leave-one-out. Т.к. если использовать k-fold, то модель мудет каждый раз недообучаться, т.к. данных и так мало. Так же это не займет много времени из-за размера выборки.
2)т.к. выборка большая, то leave-one-out будет очень долго считаться, поэтому надо использоваться k-fold, и т.к. данных много, то потеря sample.size()/k данных должно не сильно повлиять на обобщающую способность модели и будет противостоять переобучению.

Задание 2.
1) p = 2
окружности с центром в начале координат



2) ромбики
3) квадраты



Задание 3.

$$P\left(\min_{i=1..l} \rho(0, x_i) \le \rho\right) = 1 - P\left(\rho(0, x_1) > \rho, \ldots, \rho(0, x_l) > \rho\right) =$$

$$= 1 - P\left(\rho(0, x_1) > \rho\right) \cdot \ldots \cdot P\left(\rho(0, x_l) > \rho\right) =$$

НЕЗАВИСИМЫ

$$= 1 - (1 - \rho^n)^l$$

РАВНОМЕРНО

$$1 - (1 - \rho_{\frac{1}{2}}^n)^l = \frac{1}{2}$$

$$(1 - \rho_{\frac{1}{2}}^n) = \sqrt[l]{\frac{1}{2}}$$

$$\rho_{\frac{1}{2}} = \left(1 - \sqrt[l]{\frac{1}{2}}\right)^{\frac{1}{n}}$$

при $n \to \infty$ $\rho_{\frac{1}{2}} \to 1$

1) $n = 10$ $l = 500$ $\Rightarrow \rho_{\frac{1}{2}} = 0.51$

2) $n = 100$ $l = 500$ $\Rightarrow \rho_{\frac{1}{2}} = 0.936$

3) $n = 1000$ $l = 500$ $\Rightarrow \rho_{\frac{1}{2}} = 0.9934$

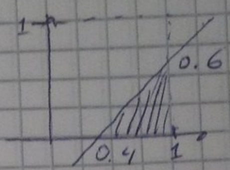
Задание 4.

1) $p = 1 \qquad (0.1, x_1) \qquad (0.5, x_2)$

$P(|x_2| + |0.5| < |x_1| + |0.1|) = P(x_2 + 0.4 < x_1) =$

$= P(x_2 < x_1 - 0.4) = \frac{1}{2} \cdot 0.6 \cdot 0.6 =$

$= 0.18$



2) $P\left((x_2^q + 0.5^q)^{\frac{1}{q}} < (x_1^q + 0.1^q)^{\frac{1}{q}} \right) =$

$$= \int_a^b \int_0^{(x_1^q + 0.1^q - 0.5^q)^{\frac{1}{q}}} dx_2\, dx_1$$

I $\underline{q = 2}$ $\Rightarrow \int_{\sqrt{0.24}}^{1} \int_0^{\sqrt{x_1^2 - 0.24}} dx_2\, dx_1 = \int_{\sqrt{0.24}}^{1} \sqrt{x_1^2 - 0.24}\, dx_1 =$

$$= \frac{1}{10} x_1 \sqrt{25x_1^2 - 6} - \frac{3}{25} \log\left(\sqrt{25x_1^2 - 6} + 5x_1\right) \Big|_{\sqrt{0.24}}^{1}$$

3) $P(\max(|x_1|, |0.1|) > \max(|x_2|, |0.5|)) =$

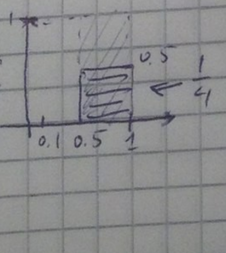
$= \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$

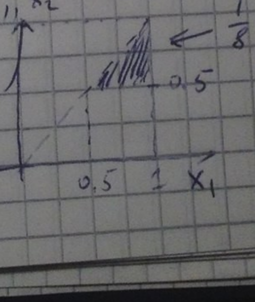
I) $x_1 \le 0.1 \Rightarrow P(0.1 > \max(x_2, 0.5)) = 0$

II) $x_1 > 0.1 \Rightarrow P(x_1 > \max(x_2, 0.5))$ разобьем событие

II') $P(x_1 > 0.1,\ x_2 \le 0.5,\ x_1 > 0.5) = \frac{1}{4}$

II'') $P(x_1 > 0.1,\ x_2 > 0.5,\ x_1 > x_2) = \frac{1}{8}$

II':


II'':


Задание 5.
“Чем больше метод машинного обучения склонен к переобучению, тем боль-
ше настраиваемых параметров у него должно быть.” Алгоритм склонен к переобучение, если он начинает подстраиваться под каждого представителя выборки и плохо перестает экстраполировать и представлять всю генеральную совакупность целиком. Это заложено в методе построение алгоритма(допускает ли он подогнанность с числом операций под конкертную выборку), а многие настраиваемые параметры могут наоборот сбособствовать предотвращению оверфитинга. Например в xgboost есть параметры регуляризации, максимальная глубина, минимальное число элементов в листе, минимальное улучшение на каждом шаге и другие.
В данном примере линейная модель строит одну кривую по всем точкам для всей выборки.
А k-NN для каждого отдельного элемента выборки определяет его класс, тем самым он более гибко описывает множество объектов.

Задание 6.
Рассмотрим алгоритм построение дерева:
1) в каждой вершине дерева мы бежим циклом длины n, т.е. количеству фичей.(ищим на какой фиче наилучшее разбиение);
2)внутри этого цикла (т.е. для каждой фичи), мы сначала сортируем значения фичи за O(l log(l))(где l размер выборки)
затем мы один раз за O(l), подсчитываем функционал качества(ошибку), отдельно храним ошибку справа от разделителя и слева.
Теперь мы должны пройтись циклом по всем уникальным отсортированным значения фичи (не длиннее l), но нам не надо еще кадый раз считать за l операций ошибку, за счет особого хранения можно не всю ее заного пересчитывать, а вычитать из правой части ошибку по длине сдвига и прибавлять к левой.
Затем выбирается точка, на которой был достигнут минимум O(l) и смотрят на другую фичу.
3) всего вершин в дереве некоторая константа, которую мы контролируем (max_depth) она не влияет на ассимптотику
Итого получили: O(n(l log(l) + l +l +l)) = O(n l log(l)

Задание 7.

1) $\exists\ |R_m| = n = n_1 + \dots + n_K$

$$\frac{1}{n} E \sum_{i=1}^{n} I(y_i \ne K_i) = E \sum_{i=1}^{K} \sum_{j=1}^{n_i} \frac{I(x_{ji} \ne i)}{n} =$$

$$= \sum_{i=1}^{K} p_{mi}\, E(I(x_{ji} \ne i)) = \sum_{i=1}^{K} p_{mi}(1 - p_{mi})$$

2) $$\sum_{k=1}^{K} p_{mk}(1-p_{mk}) = \sum_{k \ne \max_k p_{mk}}^{K} p_{mk}(1-p_{mk}) + p_{max}(1-p_{max}) \Rightarrow$$

$\Rightarrow$ Рассмотрим разницу

$$\sum_{k \ne \max_k p_{mk}} p_{mk}(1-p_{mk}) + \underbrace{p_{max}(1-p_{max}) - (1-p_{max})}_{-(1 - 2p_{max} + p_{max}^2)} =$$

$$= \sum_{k \ne \max} p_{mk}(1-p_{mk}) - (1-p_{max})^2 \ge \sum_{k \ne} p_{mk}(1-p_{max}) - (1-p_{max})^2 \ge$$

$$\ge (1-p_{max})\Big(\sum_{k \ne} p_{mk} + p_{max} - 1\Big) = (1-p_{max})\Big(\underbrace{\sum_{k=1}^{K} p_{mk}}_{=1} - 1\Big) = 0$$

≥ 0

$\Rightarrow$ преобладающий класс не хуже $a(x)$

3) $$\frac{1}{n} \sum_{i=1}^{n} \big(I(y(x_i)=k) - \bar{y}\big)^2 =$$

$$\bar{y} = \frac{\sum_{i=1}^{n} I(y(x_i)=k)}{n} = p_{mk}$$

$$I_{ik} = I(y(x_i)=k)$$

$$= \frac{1}{n} \sum_{k=1}^{K} \big(p_{mk}\cdot n - 2 p_{mk}\cdot(p_{mk} n) + n p_{mk}^2\big) =$$

$$= \sum_{k=1}^{K} (p_{mk} - p_{mk}^2) = \sum_{k=1}^{K} p_{mk}(1-p_{mk})$$