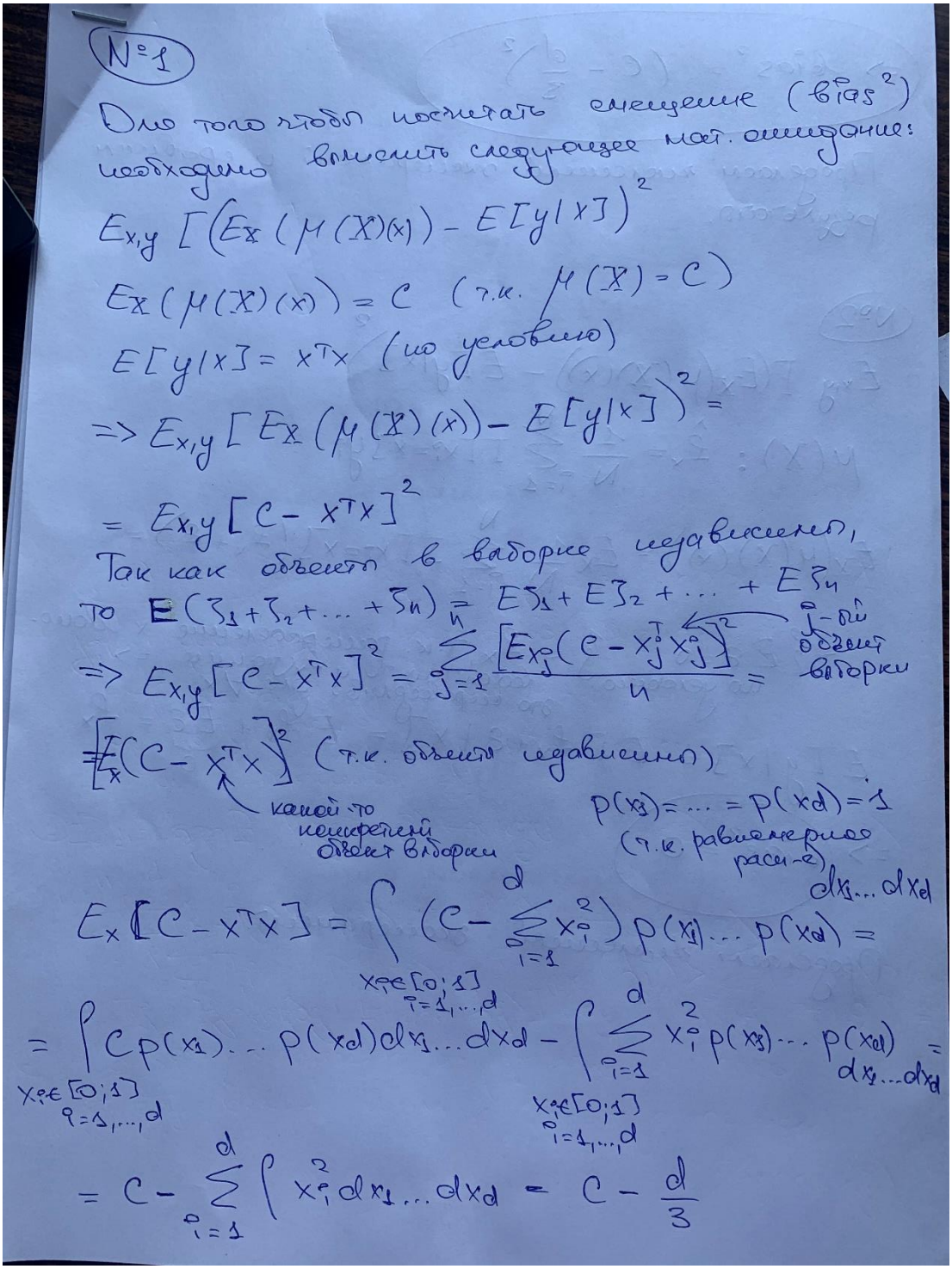
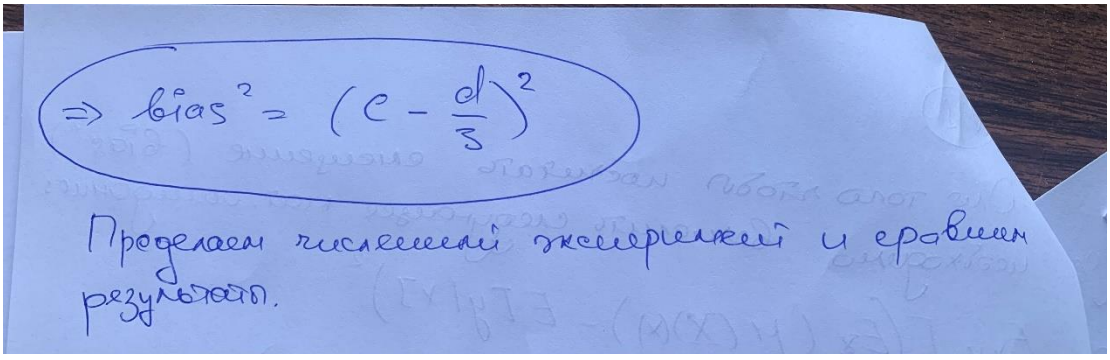
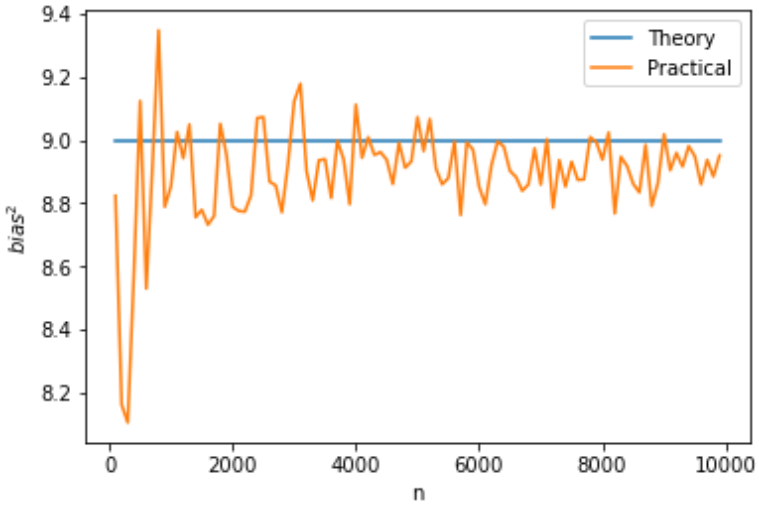
1.


№1
Для того чтобы посчитать смещение ($bias^2$) необходимо вычислить следующее мат. ожидание:
$E_{x,y}\left[\left(E_x(\mu(X)(x)) - E[y|x]\right)^2\right]$
$E_x(\mu(X)(x)) = C$ (т.к. $\mu(X) = C$)
$E[y|x] = x^T x$ (по условию)
$\Rightarrow E_{x,y}\left[E_x(\mu(X)(x)) - E[y|x]\right)^2 =$
$= E_{x,y}\left[C - x^T x\right]^2$
Так как объекты в выборке независимы,
то $E(\xi_1 + \xi_2 + \dots + \xi_n) = E\xi_1 + E\xi_2 + \dots + E\xi_n$
$\Rightarrow E_{x,y}\left[C - x^T x\right]^2 = \sum_{j=1}^{n} \frac{\left[E_{x_j}(C - x_j^T x_j^0)\right]^2}{n} =$
j-ой объект выборки
$\left[E_x(C - x^T x)\right]^2$ (т.к. объекты независимы)
какой-то конкретный объект выборки
$p(x_1) = \dots = p(x_d) = 1$
(т.к. равномерное расп-е)
$E_x\left[C - x^T x\right] = \int_{x_i \in [0;1],\ i=1,\dots,d} \left(C - \sum_{i=1}^{d} x_i^2\right) p(x_1)\dots p(x_d)\, dx_1 \dots dx_d =$
$= \int_{x_i \in [0;1],\ i=1,\dots,d} C\, p(x_1)\dots p(x_d)\, dx_1 \dots dx_d - \int_{x_i \in [0;1],\ i=1,\dots,d} \sum_{i=1}^{d} x_i^2\, p(x_1)\dots p(x_d)\, dx_1 \dots dx_d =$
$= C - \sum_{i=1}^{d} \int x_i^2\, dx_1 \dots dx_d = C - \frac{d}{3}$


$\Rightarrow bias^2 = \left(C - \frac{d}{3}\right)^2$
Проделаем численный эксперимент и сравним результаты.



Теоретический расчет и результаты эксперимента не противоречат друг другу, при увеличении размера матрицы видна тенденция стремления к теоретическому значению.

2.

№2

$$E_{x,y}\left[\left(E_X\left(\mu(X)(x)\right) - E[y|x]\right)^2\right]$$

$$\mu(X): \hat{f}_x = \frac{1}{N}\sum_{i=1}^{N} I[x_i = x]\, y_i$$

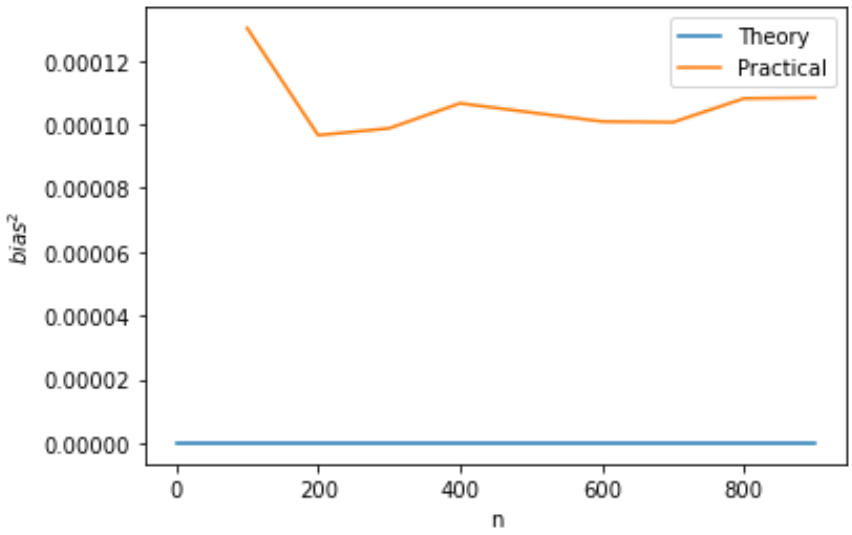
$$E_X\left(\mu(X)(x)\right) = E_X \frac{1}{N}\sum_{i=1}^{N} I[x_i = x]\left(f_x + \varepsilon\right) =$$

$= \{E\varepsilon = 0$ по условию$\} = f_x$, т.к. распределение x равномерное, то и f_x тоже, $\Rightarrow$ это справедливо для $\forall x$.

$$E[y|x] = E[f_x + \varepsilon | x] = E f_x + \underbrace{E\varepsilon}_{0} = f_x$$

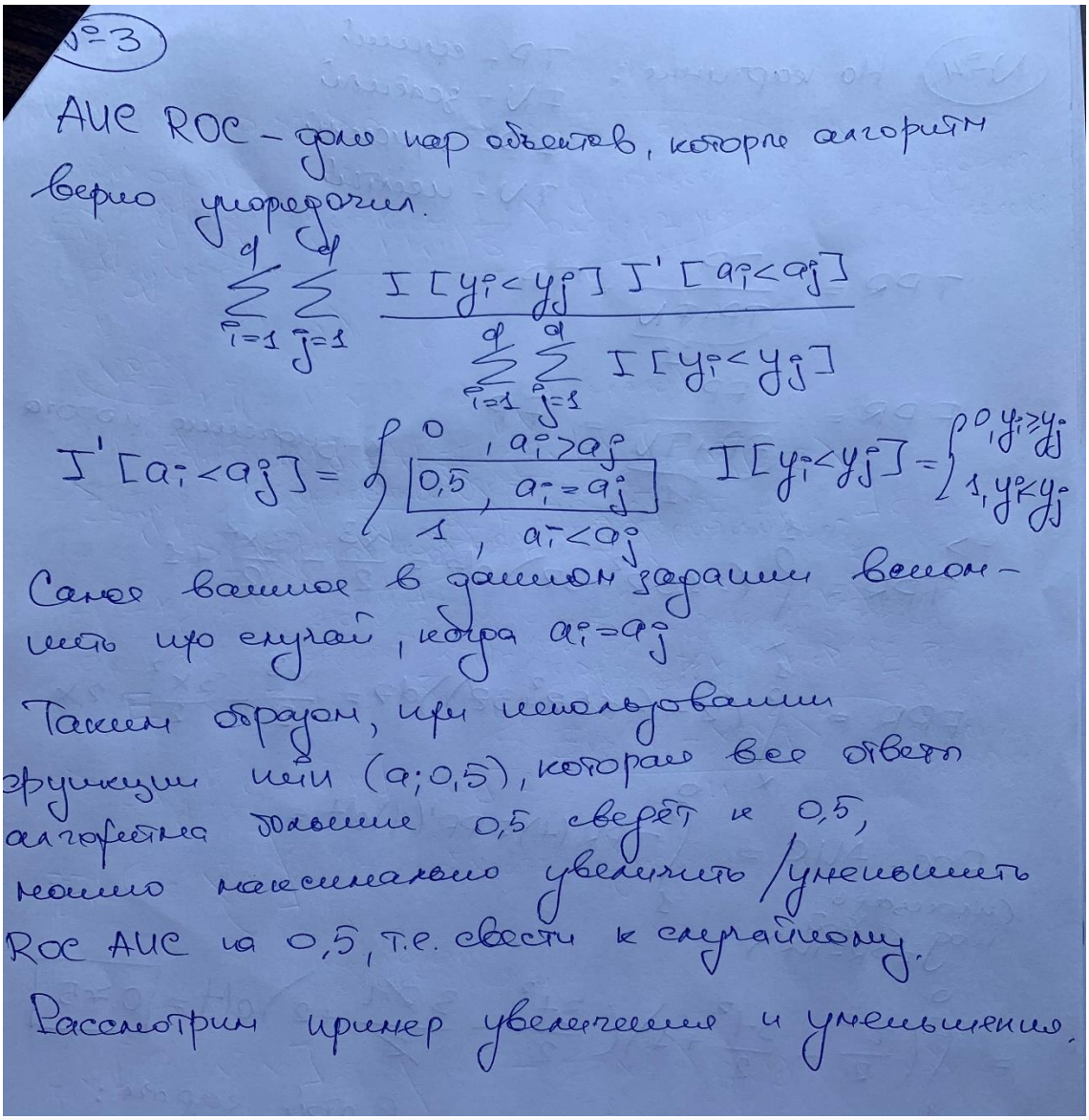
$$\Rightarrow \boxed{bias^2 = 0}$$

Проделаем численный эксперимент.



Эксперименты в среднем дают около 0.0001, что близко к теоретически рассчитанному значению 0.

3.




```
# уменьшаем на 0.5

y_true = [1, 1, 1, 0, 0, 0]
y_pred = [0.9, 0.8, 0.6, 0.5, 0.5, 0.5]

print('ROC_AUC : ', roc_auc_score(y_true, y_pred))

y_true = [1, 1, 1, 0, 0, 0]
y_pred = [0.5, 0.5, 0.5, 0.5, 0.5, 0.5]

print('ROC_AUC : ', roc_auc_score(y_true, y_pred))
```

```
ROC_AUC :  1.0
ROC_AUC :  0.5
```

```
# увеличиваем на 0.5

y_true = [0, 1, 1, 1, 1, 1]
y_pred = [0.9, 0.8, 0.6, 0.5, 0.5, 0.5]

print('ROC_AUC : ', roc_auc_score(y_true, y_pred))

y_true = [0, 1, 1, 1, 1, 1]
y_pred = [0.5, 0.5, 0.5, 0.5, 0.5, 0.5]

print('ROC_AUC : ', roc_auc_score(y_true, y_pred))
```

```
ROC_AUC :  0.0
ROC_AUC :  0.5
```

4.

№4 На картинке: TP – синий
FN – зелёный
FP – красный
TN – жёлтый

$$TPR = \frac{TP}{\underbrace{TP+FN}_{1}} = TP$$

$$FPR = \frac{FP}{\underbrace{FP+TN}_{1}} = FP$$

x – значение порога

$$\Rightarrow TPR = \int_{x}^{1} (-1{,}5z^2+3z)\,dz = \frac{x^3}{2} - \frac{3x^2}{2} + 1$$

$$FPR = \int_{x}^{1} (2-(-1{,}5z^2+3z))\,dz = \frac{x^3}{2} - \frac{3x^2}{2} + 2x$$

$$\text{Roc-AUC} = \int_{1}^{0} TPR(t)\,FPR'(t)\,dt =$$

(площадь под кривой FPR-TPR)

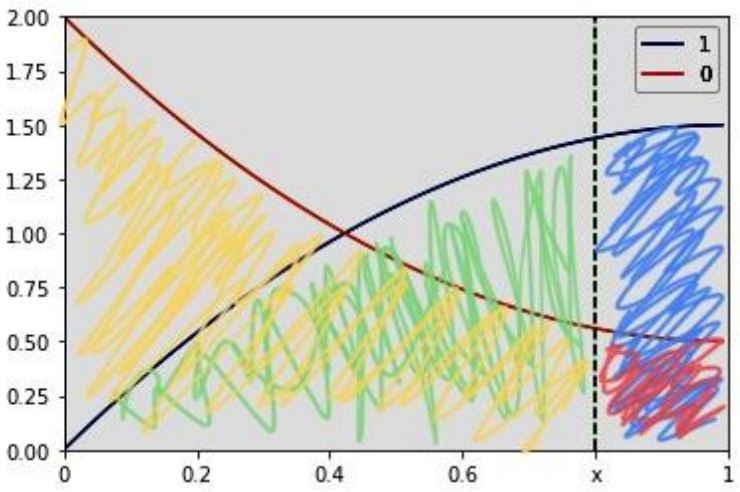
$$= \int_{1}^{0} \left(\frac{t^3}{2} - \frac{3t^2}{2} + 1\right)\left(\frac{3t^2}{2} - \frac{3 \cdot 2 \cdot t}{2} + 2\right) dt = \boxed{0{,}75}$$

Лучший порог разделения в этой задаче:

$-1{,}5z^2+3z = 2+1{,}5z^2-3z$

$z \approx 0{,}42$

В этом случае $TPR = TP \approx 0{,}77$
$FPR = FP \approx 0{,}61$



Интересно посчитать accuracy

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

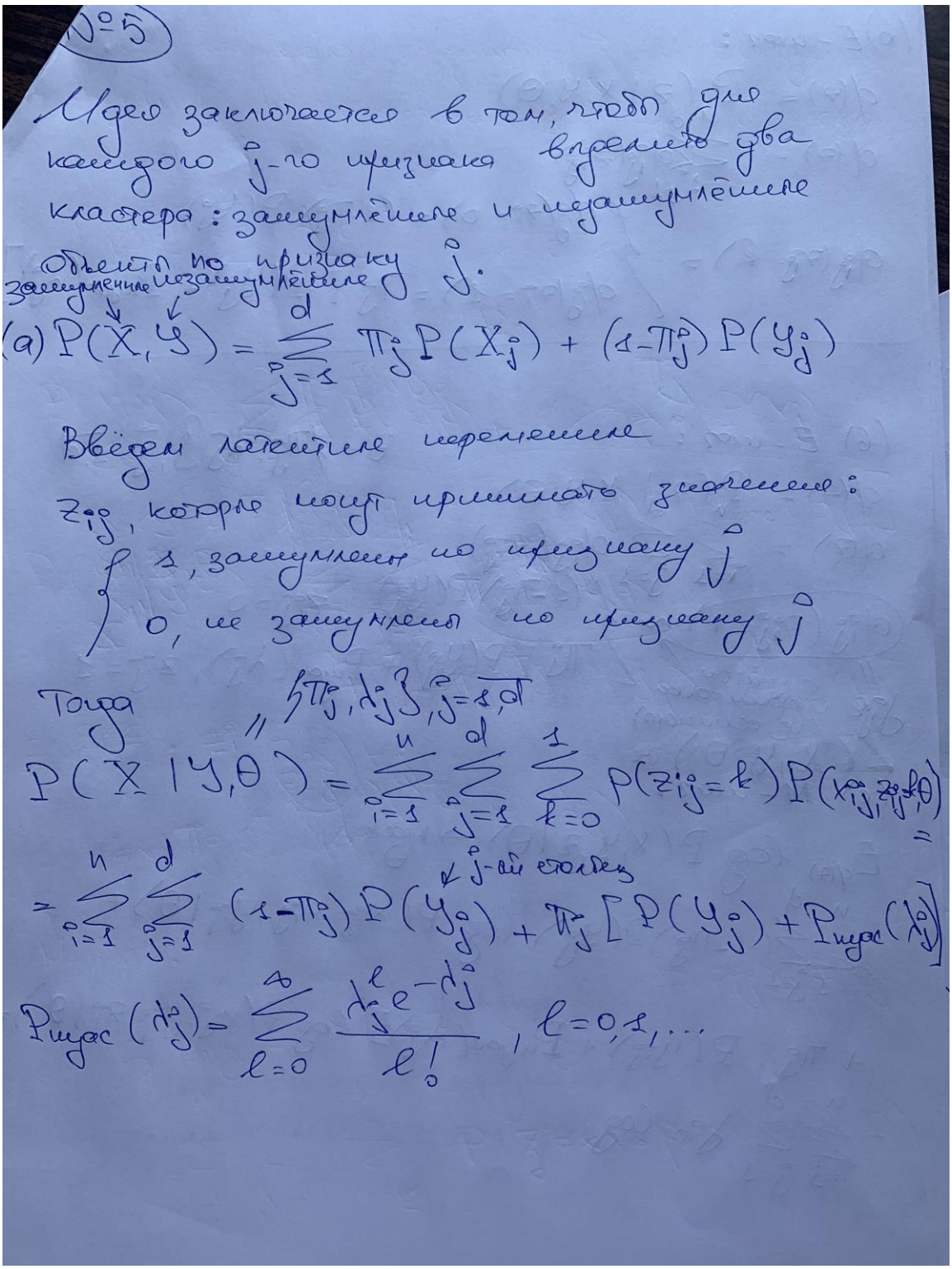
$$TN = \int_0^x (2 + 1{,}57z^2 - 3z)\,dz = \frac{x^3}{2} - \frac{3x^2}{2} + 2x$$

При $x = 0{,}42$ $TN \approx 0{,}61$, =>

accuracy $\approx 0{,}69$

То есть, ROC-AUC зависит максимально дифференцируемо.

5.



№5)

Идея заключается в том, чтобы для каждого j-го признака вводилось два кластера: зашумлённые и незашумлённые объекты по признаку j.

(а) $P(X, Y) = \sum_{j=1}^{d} \pi_j^j P(X_j^j) + (1-\pi_j^j) P(Y_j^j)$

(зашумление — незашумление)

Введём латентную переменную z_{ij}, которая может принимать значения:

$\begin{cases} 1, \text{ зашумлен по признаку } j \\ 0, \text{ не зашумлен по признаку } j \end{cases}$

Тогда $\theta = \{\pi_j, \lambda_j\}_{j=\overline{1,d}}$

$P(X \mid Y, \theta) = \sum_{i=1}^{n} \sum_{j=1}^{d} \sum_{k=0}^{1} P(z_{ij}=k) P(x_{ij}, z_{ij}=k \mid \theta) =$

$= \sum_{i=1}^{n} \sum_{j=1}^{d} (1-\pi_j) P(y_{ij}) + \pi_j [P(y_j) + P_{пуас}(\lambda_j)]$

(j-ый столбец)

$P_{пуас}(\lambda_j) = \sum_{l=0}^{\infty} \frac{\lambda_j^l e^{-\lambda_j}}{l!}, \quad l = 0, 1, \ldots$

(b) E-шаг:

$$q(z) = P_{u\,d}(z \mid X, Y, \Theta)$$

$$q(z) = \prod_{i=1}^{n}\prod_{j=1}^{d} q_{ij}(z_{ij})$$

$$q_{ij}(z_{ij} = k) = \begin{cases} q_{ij}(z_{ij} = 0) = (1 - \pi_j)\, P(y_{ij}) \\ q_{ij}(z_{ij} = 1) = \pi_j \left[P(y_{ij}) + P_{нуле}(\lambda_j) \right] \end{cases}$$

(b) E-шаг:

$$\Theta = \{\pi_j, \lambda_j\}_{j=\overline{1,d}}$$

$$q(z) = P_{u\,d}(z \mid X, Y, \Theta)$$

$$q(z) = \prod_{i=1}^{n}\prod_{j=1}^{d} q_{ij}(z_{ij})$$

$q_{ij}(z_{ij} = k)$ = (с учётом нормировки (из семинара))

$$\begin{cases} (1 - \pi_j)\, P(y_{ij}), & z_{ij} = 0 \\ \pi_j \left[P(y_{ij}) + P_{нуле}(\lambda_j) \right], & z_{ij} = 1 \end{cases}$$

(c) M-шаг:

$$E_{q(z)} \log P(X, Y, z \mid \Theta) \to \max_{\Theta}$$

$$\prod_{i=1}^{n}\prod_{j=1}^{d} (1 - \pi_j) P(y_{ij}) \log P(x_{ij}, z_{ij}, y_{ij} \mid \Theta) +$$

$$+ \pi_j \left[P(y_{ij}) + P_{нуле}(\lambda_j) \right] \log P(x_{ij}, y_{ij}, z_j \mid \Theta) =$$

$$= \prod_{i=1}^{n}\prod_{j=1}^{d} \lambda_{ij} \log$$

Eq(7) $\log P(X, y, z \mid \theta) \to \max_{\theta}$

$$\sum_{i=1}^{n} \sum_{j=1}^{d} \sum_{k=0}^{1} d_{ij}^{k} [z_{ij} = k] \log P(x_{ij}, y_{ij}, z_{ij} \mid \theta) =$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{d} d_{ij} y_{ij} \left[\log P(y_{ij}) + \log \pi_j \right] +$$

$$+ (1 - y_{ij}) \left[\log \left(P(y_{ij}) + P_{пуас}(\lambda_j) \right) + \log (1 - \pi_j) \right]$$

$$\xrightarrow[\lambda_j, \pi_j]{\max}$$

(d) Пусть на вход поступает матрица X'.
Необходимо для каждого признака j
определить зашумлён он или нет.
То есть, для $\forall j$ определить значение
скрытой переменной z!

① Для этого надо сравнить $\phi(z=0)$ и $\phi(z=1)$
признак j зашумлён, если

$$\sum_{i=1}^{n} \pi_j \left[P(y_{ij}) + P_{пуас}(\lambda_j) \right] > \sum_{i=1}^{n} (1 - \pi_j) P(y_{ij})$$

② Из зашумлённых признаков вместо
Пуассоновский шум с распределением
$P_{пуас}(\lambda_j)$