Общее: Каждая задача оценивается целым числом баллов от 0 до 6. Все результаты масштабируются в процентах от второй наилучшей работы.

1. Ёжику досталась обучающая выборка для задачи классификации:
   целевая переменная $y = (1, -1, -1, 1, 1)$ и предиктор $x = (1, 2, 3, 4, 5)$.

   Медвежонок выбирает одно наблюдение из пяти равновероятно, обозначим с помощью Y и X полученные случайные значения целевой переменной и предиктора. Ёжик может задать Медвежонку один вопрос вида «правда ли, что X больше константы c?».

   Какую c надо выбрать Ёжику, чтобы минимизировать $H(Y|Q)$, где Q — это ответ Медвежонка?

   Подсказка: если $h(t) = -t \ln t$, то $h(1/2) = 0.347$, $h(1/3) = 0.366$, $h(2/3) = 0.270$, $h(3/4) = 0.216$, $h(2/5) = 0.367$, впрочем, задачу можно полностью решить и без этих цифр :)

   Решение:

   $$H(Y|Q) = \mathbb{P}(Q = \text{да})H(Y|Q = \text{да}) + \mathbb{P}(Q = \text{нет})H(Y|Q = \text{нет})$$

   Обозначим $a(p)$ — энтропию дискретной величины, принимающей значения 1 и (-1) с вероятностями p и $1 - p$, сразу замечаем, что $a(0) = a(1) = 0$, и чем ближе p к $1/2$, тем больше $a(p)$, $a(p) = a(1 - p)$.

   Делаем быстрый перебор:

   $c = 0$ (всё в одну кучу):
   $$H(Y|Q) = a(0.4)$$
   $c = 1.5$:
   $$H(Y|Q) = 0.2a(1) + 0.8a(1/2) = 0.8a(1/2)$$
   $c = 2.5$:
   $$H(Y|Q) = 0.4a(1/2) + 0.6a(1/3)$$
   $c = 3.5$:
   $$H(Y|Q) = 0.6a(1/3) + 0.4a(1) = 0.6a(1/3)$$
   $c = 4.5$:
   $$H(Y|Q) = 0.8a(1/2) + 0.2a(1) = 0.8a(1/2)$$

   Замечаем, что $a(1/3) < a(0.4) < a(1/2)$, можно устно, можно и числа подставить.

   Следовательно, $0.6a(1/3)$ меньше, чем $a(0.4)$, $0.8a(1/2)$ или $0.4a(1/2) + 0.6a(1/3)$.

   Все $c \in [3; 4)$ подойдут.

   Верно применена формула для условной энтропии хотя бы один раз: +3 балла.

   Аккуратный перебор всех случаев: + 3 балла.

   Если угадано $c = 3$ с неверным объяснением: +1 за удачливость.

2. Ежу понятно, что математическое ожидание первой производной лог-функции правдоподобия тождественно равно нулю. Производная нуля равна нулю. Для хорошей функции правдопобия производная от ожидания первой производной равна ожиданию второй производной. Ожидание второй производной лог-функции правдоподобия со знаком минус называется информацией Фишера. Следовательно, информация Фишера всегда равна нулю.

Найдите качественную ошибку в этом рассуждении.

Решение:

Ожидание производной лог-правдоподобия зависит от двух аргументов: истинного параметра, по которому считается ожидание, и аргумента лог функция правдоподобия, от которой считается ожидание.

$$g(\theta, \theta^T) = \int_{\mathbb{R}^n} f(y \mid \theta^T) \frac{\partial}{\partial \theta} \ln f(y \mid \theta) dy.$$

Производная равна нулю только в точке, где аргумент функции лог-правдоподобия совпадает с истинным параметром,

$$g(\theta^T, \theta^T) = 0$$

Равенство нулю производной в точке не означает равенства нулю следующей производной. Всё.

$$\frac{\partial}{\partial \theta} g(\theta, \theta^T) \neq 0 \text{ в точке } \theta = \theta^T.$$

Достаточно текстового объяснения без формул.

Если угадано, в каком логическом переходе ошибка, но не объяснено, какая именно: +1.

Упомянута регулярность: +1.

Часто говорили, что у функции правдоподобия оптимум там, где производная равна нулю, но в задаче речь об ожидании правдоподобия.

3. Хорошо обученная свинья тратит на поиск одного трюфеля экспоненциальное время с ожиданием μ минут. Поиск различных трюфелей независим. Жерар Депардье, к сожалению, замерял время поиска 100 трюфелей только тремя диапазонами: от 0 до 10 минут (20 трюфелей), от 10 до 20 минут (50 трюфелей), более 20 минут (30 трюфелей).

   а) Постройте 95% асимптотический доверительный интервал для μ.
   б) С помощью LM-теста проверьте гипотезу о том, что $\mu = 10$ против альтернативной гипотезы о неравенстве для уровня значимости 5%.

Табличное: правые 5% критические значения: $\chi^2_1 = 3.84$, $\chi^2_2 = 5.99$, $\chi^2_3 = 7.81$, $\chi^2_4 = 9.49$, функция плотности экспоненциального распределения имеет вид $\lambda \exp(-\lambda x)$.

Решение:

Для экспоненциального распределения вероятность попасть от a до b равна $\exp(-a/\mu) - \exp(-b/\mu)$.

В нашем случае: $\mathbb{P}(Y_i < 10) = (1 - \exp(-10/\mu)) = 1 - t$, $\mathbb{P}(Y_i \in [10; 20]) = (\exp(-20/\mu) - \exp(-10/\mu)) = t(1-t)$, $\mathbb{P}(Y_i > 20) = \exp(-20/\mu) = t^2$.

Функция правдоподобия равна

$$L = \frac{100!}{20!30!50!} (1-t)^{20} (t(1-t))^{50} (t^2)^{30} = ct^{110}(1-t)^{70}$$

Экстремум равен $\hat{t} = 110/180 = 11/18$, $-10/\hat{\mu} = \ln 11 - \ln 18$,

Выписана функция правдоподобия: 1 балл.

Найдена оценка: 1 балл.

Найдена стандартная ошибка: 1 балл.

Найден доверительный интервал: 1 балла.

Проведен LM тест: 2 балла.

4. Ёж проверяет всего две гипотезы H_0^i одновременно, одна из которых верна, а вторая — нет. Для неверной гипотезы P-значение распределено равномерно на $[0; 0.5]$.

   Ёж сортирует P-значения по возрастанию, $p_{(1)} \leq p_{(2)}$. Затем сравнивает $p_{(1)}$ с константой 0.05. Если $p_{(1)} \geq 0.05$, то обе $H_0^{(i)}$ не отвергаются и Ёж заканчивает работу. Если $p_{(1)} < 0.05$, то $H_0^{(1)}$ отвергается и Ёж сравнивает $p_{(2)}$ с константой b. Если $p_{(2)} < b$, то $H_0^{(2)}$ отвергается, иначе $H_0^{(2)}$ не отвергается.

   Постройте график зависимости FDR (false discovery rate) от $b \in [0.05; 1]$.

   Пусть N_R — число отвергнутых нулевых гипотез, а N_{WR} — число ошибочно отвергнутых нулевых гипотез.

   У нас всего-то три случая:

$$FDR = \mathbb{E}\left(\frac{N_{WR}}{\max\{N_R, 1\}}\right) = \sum_{i=0}^{2} \mathbb{E}\left(\frac{N_{WR}}{\max\{N_R, 1\}} \cdot I(N_R = i)\right)$$

   Находим каждое слагаемое. Первое:

$$\mathbb{E}\left(\frac{N_{WR}}{\max\{N_R, 1\}} \cdot I(N_R = 2)\right) = 0.5\,\mathbb{P}(N_R = 2)$$

   Находим вероятность того, что обе гипотезы отвергнуты. Это вероятность того, что оба P-значения меньше b и хотя бы одно из них меньше 0.05. Можно закрасить соответствующую фигуру на прямоугольнике, где совместная плотность больше нуля, и помножить площадь на плотность, получится:

$$\mathbb{P}(N_R = 2) = \text{площадь} \cdot 2 = (2 \cdot 0.05b - 0.05^2) \cdot 2.$$

   Второе:

$$\mathbb{E}\left(\frac{N_{WR}}{\max\{N_R, 1\}} \cdot I(N_R = 1)\right) = \mathbb{P}(N_{WR} = 1, N_R = 1)$$

   Находим вероятность того, что отвергнута только та гипотеза, что была верна.

$$\mathbb{P}(N_{WR} = 1, N_R = 1) = \mathbb{P}(p_T < 0.05, p_F > b) = \mathbb{P}(p_T < 0.05)\,\mathbb{P}(p_F > b) = 0.05 \cdot (1 - 2b).$$

   Третье:

$$\mathbb{E}\left(\frac{N_{WR}}{\max\{N_R, 1\}} \cdot I(N_R = 0)\right) = 0$$

   Выписана FDR в верном виде применительно к этой задаче (видно, что будет три случая): +3

   Рассмотрен каждый случай: +1 за случай.

5. Пчёлы бывают правильные ($b_i = \text{good}$) и неправильные ($b_i = \text{bad}$). Из одного дупла правильных пчёл можно извлечь случайное равномерное количество мёда, $(y_i \mid b_i = \text{good}) \sim U[0; a]$, где параметр a не известен и $a > 1$. Для одного дупла неправильных $(y_i \mid b_i = \text{bad}) \sim U[0; 1]$. Имеется n независимых наблюдений. Неизвестную вероятность того, что в дупле водятся правильные пчёлы, обозначим буквой π.

   Явно выпишите целевую функцию для M-шага EM-алгоритма в этой задаче, поясните по каким параметрам она оптимизируется и смысл остальных параметров.

   Решение:

   Обозначим $g_i = \mathbb{P}(b_i = \text{good} \mid a_{old}, \pi_{old})$. В задаче не требовалось находить эту функцию, но вот она:

$$
g_i = \begin{cases} 1, \text{ если } y_i > 1; \\ \frac{\pi_{old}/a_{old}}{\pi_{old}/a_{old} + (1-\pi_{old})\cdot 1}, \text{ если } y_i \leq 1. \end{cases}
$$

   Целевая функция равна:

$$
Q(a, \pi \mid a_{old}, \pi_{old}) = \begin{cases} \sum_{i=1}^n g_i \ln(\pi/a) + (1-g_i)\ln(1-\pi), \text{ если } y_i \leq 1 \text{ при всех } g_i < 1; \\ -\infty, \text{ иначе.} \end{cases}
$$

   Максимизируем по a, π.

   Корректно выписано в общем виде и при этом разделены тета old от тета: 3 балла.

   Явно указано по каким переменным идет оптимизация: +1 балл.

   Явно выписаны функции плотности для данной задачи: +2 балла.