Домашнее задание по оценкам обобщающей способности 1.
ММП, весна 2013
8 марта

1 Неравенства концентрации

1. Рассмотрим сумму n независимых случайных величин ξ_i Бернулли с одинаковыми параметрами $\mathbb{E}[\xi_i] = p$ — то есть биномиальную случайную величину с параметрами (n, p): $S_n = \sum_{i=1}^n \xi_n$. Выпишите верхнюю оценку вероятности отклонений случайной величины $\frac{S_n}{n}$ от ее математического ожидания в большую сторону $\mathbb{P}\{S_n/n - \mathbb{E}[S_n/n] \geqslant t\}$ с помощью неравенства Хевдинга.

2. Теперь рассмотрим сумму n независимых случайных величин ξ_i, равномерно распределенных на отрезке $[0, 1]$: $S_n = \sum_{i=1}^n \xi_n$. Выпишите верхнюю оценку вероятности отклонений случайной величины $\frac{S_n}{n}$ от ее математического ожидания в большую сторону $\mathbb{P}\{S_n/n - \mathbb{E}[S_n/n] \geqslant t\}$ с помощью неравенства Хевдинга.

3. Проведите следующий численный эксперимент. Постройте функции эмпирических распределений двух случайных величин $S_n/n - \mathbb{E}[S_n/n]$ из прошлых задач. (*Для этого зафиксируйте размер выборки n, например $n = 100$, и какое-нибудь достаточно большое N, например $N = 10^5$, и сгенерируйте N независимых наблюдений случайной величины $S_n/n - \mathbb{E}[S_n/n]$. Проделайте это два раза — для суммы Бернулли и суммы равномерных случайных величин. Затем по полученным выборкам можно эмпирически изучить поведение хвоста $\mathbb{P}\{S_n/n - \mathbb{E}[S_n/n] \geqslant t\}$: ввести сетку по t и для каждого узла сетки t_i считать долю наблюдений выборки, превосходящих t_i.*) Сравните два полученных хвоста. Что вы можете о них сказать?

4. На том же рисунке изобразите верхние оценки, полученные в первых двух задачах с помощью неравенства Хевдинга. Завышены ли они? Подумайте, почему наблюдается полученная вами картина? Может быть, оценка Хевдинга не учитывает чего-то важного?

5. Рассмотрим множество A ограниченных случайных величин, с вероятностью 1 принимающих значения из интервала $[0, 1]$. Каково максимальное значение дисперсии таких случайных величин $\sup_{\xi \in A} \mathrm{Var}[\xi]$? Достигается ли оно на какой-то конкретной случайной величине, и если да, то на какой?

6. Вспомни, как мы доказывали неравенство Хевдинга. Главным шагом было использование леммы Хевдинга — верхней оценки для преобразования Лапласа ограниченной отрезком $[a, b]$ случайной величины с нулевым мат. ожиданием: $\mathbb{E}[e^{s\xi}] \leqslant e^{s^2(b-a)^2/8}$. Что мы получим, если вместо леммы Хевдинга мы будем использовать другой результат, справедливый для случайных величин ξ с нулевым математическим ожиданием $\mathbb{E}[\xi] = 0$, на этот раз ограниченных лишь сверху $\xi < 1$:
$$\mathbb{E}[e^{s\xi}] \leqslant \exp\big((e^s - s - 1)\mathbb{E}[\xi^2]\big)?$$
   Докажите с помощью этого неравенства следующий результат, известный как *неравенство Беннета*:

   Теорема 1.1 (Неравенство Беннета) *Пусть $\xi_1, \ldots, \xi_n$ — последовательность независимых случайных величин с нулевыми мат. ожиданиями, таких что с вероятностью 1 $\xi_i \leqslant 1$. Обозначим $\sigma^2 = \frac{1}{n}\sum_{i=1}^n \mathrm{Var}[\xi_i]$.*

   Тогда для любого $t > 0$ справедливо
$$\mathbb{P}\left\{\sum_{i=1}^n \xi_i > t\right\} \leqslant \exp\left(-n\sigma^2 h\left(\frac{t}{n\sigma^2}\right)\right),$$
   где $h(u) = (1 + u)\log(1 + u) - u$ для $u \geqslant 0$.

7. С учетом неравенства $h(u) \geqslant u^2/(2 + 2u/3)$, $u \geqslant 0$ (проверьте его справедливость), получите следующее *неравенство Бернштейна* из неравенства Беннета:

   Теорема 1.2 (Неравенство Бернштейна) *В тех же условиях, что и неравенство Беннета, справедливо:*
$$\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^n \xi_i > t\right\} \leqslant \exp\left(-\frac{nt^2}{2(\sigma^2 + t/3)}\right).$$

8. Отлично! Мы получили неравенства концентрации, учитывающие дисперсию случайной величины, которые в отличие от неравенство Чебышева с ростом t убывают экспоненциально. Изобразите на том же рисунке, что и прошлые эксперименты, оценки Бернштейна для двух наших случайных величин. Улучшения наблюдаются?

2 Размерность Вапника–Червоненкиса

1. Вычислите размерности Вапника–Червоненкиса для следующих семейств классификаторов:
$$\mathcal{G}_{\mathrm{hs}} = \{g_{\boldsymbol{w},w_0}(\boldsymbol{x}) = \mathrm{sgn}(\langle \boldsymbol{x}, \boldsymbol{w}\rangle + w_0), \boldsymbol{w} \in \mathbb{R}^n, w_0 \in \mathbb{R}\}, \quad \mathbb{X} = \mathbb{R}^n, \mathbb{Y} = \{-1, +1\};$$
$$\mathcal{G}_{\sin} = \{g_t(x) = \mathrm{sgn}\big(\sin(tx)\big), t \in \mathbb{R}\}, \quad \mathbb{X} = \mathbb{R}, \mathbb{Y} = \{-1, +1\};$$
$$\mathcal{G}_{\mathrm{int}} = \{g_{a,b}(x) = x \in [a, b], a, b \in \mathbb{R}\}, \quad \mathbb{X} = \mathbb{R}, \mathbb{Y} = \{0, 1\}.$$

Подсказка: В случае $\mathcal{G}_{\text{hs}}$ для начала постройте нижнюю оценку $n+1$, указав пример множества из $n+1$ точек, разбивающихся полупространствами на всевозможные 2^{n+1} подмножеств. Затем вам надо построить строгую верхнюю оценку, доказав что любое множество из $n+2$ точек не может быть разбито всевозможными способами. Для этого воспользуйтесь следующей *теоремой Радона*:

Теорема 2.1 (Теорема Радона) *Рассмотрим множество S из $n+2$ точек в пространстве $\mathbb{R}^n$. Тогда множество S может быть разбито на два* **не пересекающихся** *подмножества $S = S_1 \cup S_2$, чьи выпуклые оболочки пересекаются.*