Домашнее задание №2

Задание 1

Количество баллов: 1

Вычислите $\mathrm{KL}(\mathcal{N}(\mu_0, \sigma_0^2) \mid\mid \mathcal{N}(\mu_1, \sigma_1^2))$, где $\mathrm{KL}(P \mid\mid Q) = \mathbb{E}_P[\log \frac{P}{Q}]$ – дивергенция Кульбака-Лейблера между двумя распределениями P и Q.

Задание 2

Количество баллов: 1.5

Аппроксимируйте растянутую функцию распределения стандартного нормального закона $\Phi(\lambda x)$ для некоторого λ с помощью сигмоиды $\sigma(x)$.

1. В ответ укажите значение $\tilde{\lambda}$, при котором производные функций совпадают в нуле.
2. Найдите численную оценку нормы $\left\|\Phi(\tilde{\lambda}\cdot) - \sigma\right\|_\infty = \sup\limits_{x\in\mathbb{R}} |\Phi(\tilde{\lambda}x) - \sigma(x)|$ с точностью 10^{-6}.
3. Нарисовать на одном графике функции $\Phi(\tilde{\lambda}x)$ и $\sigma(x)$ в диапазоне $-6 \leq x \leq 6$. Подписать оси `x/y`, указать легенду, $\Phi(\tilde{\lambda}x)$ нарисовать синим цветом, а $\sigma(x)$ оранжевым.

Примечание.

1. Здесь

$$\begin{aligned}\Phi(x) &= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp{(-t^2/2)} dt, \\ \sigma(x) &= \frac{1}{1 + \exp(-x)}.\end{aligned}$$

2. Численную оценку можно получить в Python (разрешается использовать `numpy`, `scipy` и другие библиотеки), Wolfram Matematica или Wolfram Alpha с решением в виде скриншота или файла типа `.py` / `.ipynb` / `.nb`. График также прикрепить к работе.

Задание 3

Количество баллов: 2.5

Пусть после наблюдения некоторых данных $\mathcal{D} = (\boldsymbol{X}, \boldsymbol{y})$ апостериорное распределение $p(w|\mathcal{D})$ имеет вид

$$p(w|\mathcal{D}) = 0.85\mathcal{N}(w|0, 1) + 0.15\mathcal{N}(w|4, 0.1^2),$$

где $\mathcal{N}(\cdot \mid \mu, \sigma^2)$ – плотность нормального распределения с параметрами математического ожидания μ и дисперсией σ^2, вычисленная в заданной точке.

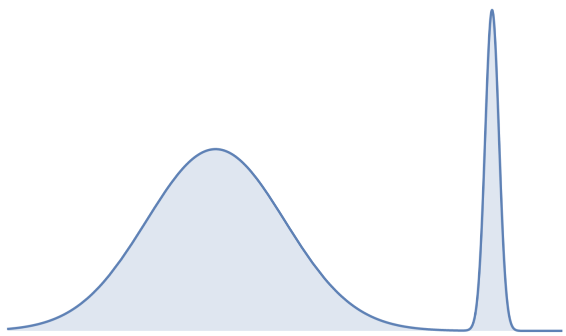

Рис. 1: Распределение $p(w|\mathcal{D})$

Предположим, что вероятностная модель имеет вид

$$p(y|x, w) = \mathcal{N}(y|xw, \sigma^2).$$

1. Проверьте, что $p(w|\mathcal{D})$ действительно задает плотность вероятностей.

2.
   - Для каждой моды w_i^{MAP} распределения $p(w|\mathcal{D})$ постройте Laplace approximation $\pi_i(w)$.
   - Для каждой аппроксимации из предыдущего пункта найдите приближенное предсказательное распределение

   $$p(y|x, \mathcal{D}) = \int p(y|x, w)p(w|\mathcal{D})dw \approx \int p(y|x, w)\pi_i(w)dw, \quad \forall i$$

   - Нарисуйте на одном полотне (`figure`) графики настоящего предсказательного распределение $p(y|x, \mathcal{D})$ (синим, посчитать его численно) и двух приближенных предсказательных распределений (оранжевым и зеленым).
   Сделать это для случаев $(x, \sigma) = (1, 1)$ и $(x, \sigma) = (2, 1)$. Подписать полотна как `$p(y | x=`x_i`, D})$`.

3. Постройте оценку $p(y|x, \mathcal{D})$, аппроксимировав апостериорное распределение распределением $q_{\boldsymbol{\theta}}(\cdot) = \mathcal{N}(\cdot|\mu, \nu^2), \quad \boldsymbol{\theta} = (\mu, \nu)$, т.ч.

   $$\mathrm{KL}(p(\cdot|\mathcal{D}) \mid\mid q_{\boldsymbol{\theta}}(\cdot)) \to \min_{\boldsymbol{\theta}}.$$

   - Укажите явный вид распределения $q_{\boldsymbol{\theta}}$.
   - Найдите аналитически приближенное предсказательное распределение $\int p(y|x, w)q_{\boldsymbol{\theta}}(w)dw$.

- Нарисуйте на одном полотне график настоящего предсказательного распределения $p(y|x, \mathcal{D})$ (синим, посчитать его численно) и его приближенной версии (оранжевым).
  Сделать это для случаев $(x, \sigma) = (1, 1)$ и $(x, \sigma) = (2, 1)$. Подписать полотна как `KL-based $p(y | x=`x_i`, D})$`.

Задание 4

Количество баллов: 1 + *[0...10]

Предположим, что у вас есть выборка $\mathcal{D} = \{(t, y_t)\}_{t=0}^{19} \subset \mathbb{N}_0 \times \mathbb{Z}$ (см. материалы к дз), которую удобно представить в виде следующего графика.

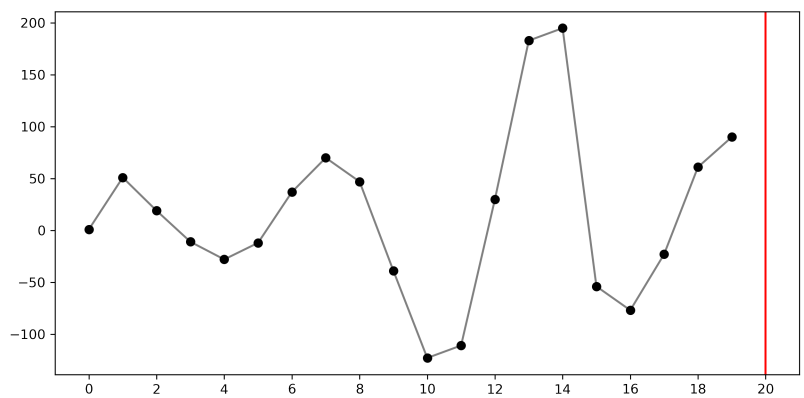

Рис. 2: Выборка $\mathcal{D} = \{(t, y_t)\}_{t=0}^{19}$

Но что это за красная линия на графике? Это область $t = 20$, для которой надо будет сделать предсказание $y = y_{20}$. Сделайте это несколькими способами:

1. Предскажите y как выборочное математическое ожидание выборки.
2. Оцените распределение $p(y|\mathcal{D})$ через метод максимального правдоподобия (т.е. аппроксимацией $p(\mathbf{w}|\mathcal{D}) \approx \delta(\mathbf{w} - \mathbf{w}_{\mathrm{ML}})$) в случае, когда модель имеет вид

   - линейной функции, т.е.

     $$y_t = w_0 + tw_1 + \varepsilon_t,$$

     где (ε_t) - гауссовский белый шум с дисперсией $\sigma^2 = 1$. Для обучения используйте последние 2 точки данных.

- квадратичной функции, т.е.

$$y_t = w_0 + tw_1 + t^2 w_2 + \varepsilon_t,$$

где (ε_t) - гауссовский белый шум с дисперсией $\sigma^2 = 1$. Для обучения используйте последние 3 точки данных.

Выпишите значения коэффициентов оценки максимального правдоподобия и явный вид приближенного предсказательного распределения.

3. * Придумайте свою (байесовскую) модель, с помощью которой сделайте точечную оценку величины y или его предсказательного распределения $p(y|\mathcal{D})$. Допустим, настоящее значение y_{20} равно y, а Ваше предсказанное равно $\hat{y}$ (в случае распределения берется мода). Тогда за этот пункт будет засчитано

$$\mathtt{rnd}\left(\frac{300}{|y-\hat{y}|+15} - 10\right)$$

баллов, где $\mathtt{rnd}(\cdot)$ округляет результат до ближайшей десятой. Это задание необязательно.

Задание 5

Количество баллов: 2

Пусть задана выборка $\{(\mathbf{x}_i, t_i)\}_{i=1}^n \subset \mathbb{R}^d \times \mathbb{R}$. Пусть модель распределения наблюдаемых имеет вид

$$t_i = \boldsymbol{\theta}^\top \mathbf{x}_i + \eta_i,$$

где $\eta_i \sim \mathcal{N}(0, \sigma^2)$, причем $\mathbb{E}[\eta_i \eta_j] = \sigma^2 \delta_{ij}$. Пусть для удобства матрица $\mathbf{X}$ – это матрица, составленная из строк $\mathbf{x}_i^\top$. Предположим, что $\mathbf{X}^\top \mathbf{X} = I$.

Выразите моду апостериорного распределения $\boldsymbol{\theta}_{\mathrm{MP}}$ для априорного распределения $p(\boldsymbol{\theta}|\lambda) = \prod_{i=1}^n \frac{\lambda}{2} \exp(-\lambda|\theta_i|)$ через оценку $\boldsymbol{\theta}_{\mathrm{ML}}$.

Задание 6

Количество баллов: 2

Пусть у нас есть выборка элементов из двух классов: $\mathbf{X} = \{(\mathbf{x}_n, 0)\}_n \subset \mathbb{R}^d \times \{0\}$ и $\mathbf{Y} = \{(\mathbf{y}_n, 1)\}_n \subset \mathbb{R}^d \times \{1\}$.
По определению, два множества $\mathbf{X}$ и $\mathbf{Y}$ называются *линейно разделимыми*, если существует вектор $\mathbf{w} \in \mathbb{R}^d$ и скаляр $w_0 \in \mathbb{R}$, т.ч. $\mathbf{w}^\top \mathbf{x}_n + w_0 > 0 \ \forall \mathbf{x}_n$ и $\mathbf{w}^\top \mathbf{y}_n + w_0 < 0 \ \forall \mathbf{y}_n$.
Докажите, что если выпуклые оболочки точек из $\mathbf{X}$ и $\mathbf{Y}$ пересекаются, то эти множества не могут быть линейно разделимы. И наоборот, если они линейно разделимы, то их выпуклые оболочки не пересекаются.

Задание 7

Количество баллов: 3

Для заданной функции ошибки $L : \mathbb{R}^2 \to \mathbb{R}_+$ и заданном совместном распределении $p(x, t)$ определен функционал $\mathcal{F}_L$, такой что

$$\mathcal{F}_L[y] = \mathbb{E}_{x,t}[L(t, y(x))] = \iint L(t, y(x))p(x, t)dxdt.$$

Найдите функцию $y = y(x)$, на которой достигается минимум функционала $\mathcal{F}_L$ при $L(t, \hat{t}) = |t - \hat{t}|$.