

# Задание 4 по курсу «Байесовский выбор моделей»

## Общая информация

- Время сдачи задания: 14е декабря, 21:00 по Москве;
- Максимальная базовая оценка за задание 100 баллов, так что при желании можно выполнять не всё;
- Оценка автора наилучшей работы удваивается (с учетом баллов сверх 100), но не более, чем до 250 баллов;
- Вопросы и само задание принимаются по почте: aduenko@phystech.edu;
- Тема письма: вопрос по заданию #4 или решение задания #4;
- Опоздание на неделю снижает оценку в 2 раза, опоздание на час на  $0.5^{1/(7 \cdot 24)} = 0.41\%$ ;
- Работы опоздавших не участвуют в конкурсе на лучшую работу;
- Задание не принимается после его разбора и / или после объявления об этом.

**Задача (байесовский метод главных компонент).** Рассмотрим вероятностную модель метода главных компонент, считая, что для каждого объекта  $\mathbf{x}_i \in \mathbb{R}^n$  существует описание  $\mathbf{z}_i \in \mathbb{R}^d$  в признаковом пространстве меньшей размерности, причем  $\mathbf{x}_i = \mathbf{W}\mathbf{z}_i + \boldsymbol{\mu} + \boldsymbol{\varepsilon}$ , где  $\boldsymbol{\mu} \in \mathbb{R}^n$  – есть некоторое смещение (на случай нецентрированности признаков), а  $\boldsymbol{\varepsilon} \in \mathbb{R}^n$  – шумовой вектор.

Пусть имеется выборка  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_m]$  независимых объектов. Пусть

$$p(\mathbf{z}_i) = N(\mathbf{z}_i | \mathbf{0}, \mathbf{I}), \quad p(\boldsymbol{\varepsilon}) = N(\boldsymbol{\varepsilon} | \mathbf{0}, \sigma^2 \mathbf{I}).$$

Считая  $\mathbf{W}$ ,  $\sigma^2$ ,  $\boldsymbol{\mu}$  – неизвестными параметрами задачи, а  $d$  фиксированным

а) выписать  $p(\mathbf{X}, \mathbf{Z} | \mathbf{W}, \boldsymbol{\mu}, \sigma)$  (3 балла);

б) найти  $p(\mathbf{X} | \mathbf{W}, \boldsymbol{\mu}, \sigma)$  (3 балла);

в) с помощью EM-алгоритма решить задачу нахождения наиболее правдоподобных оценок  $\mathbf{W}$ ,  $\boldsymbol{\mu}$ ,  $\sigma$ , то есть решить задачу

$$p(\mathbf{X} | \mathbf{W}, \boldsymbol{\mu}, \sigma) \rightarrow \max_{\mathbf{W}, \boldsymbol{\mu}, \sigma},$$

получив итеративные формулы пересчета для E и M шагов (25 баллов). Каково апостериорное распределение  $p(\mathbf{z}_i | \mathbf{x}_i, \mathbf{W}, \sigma, \boldsymbol{\mu})$ ? (10 баллов) Как изменить вероятностную модель, чтобы учесть, что в данных есть пропуски? (10 баллов)

г) сгенерировать признаковую матрицу  $\mathbf{X} \in \mathbb{R}^{m \times n}$ ,  $m = 1000$ ,  $n = 10$  для  $d = 2$  путем генерации  $\mathbf{Z} \in \mathbb{R}^{m \times d}$  поэлементно независимо из  $N(0, 1)$  и выполнения преобразования  $\mathbf{X} = \mathbf{Z}\mathbf{W}^\top + \boldsymbol{\varepsilon}$  для  $\boldsymbol{\varepsilon} \sim N(\boldsymbol{\varepsilon} | \mathbf{0}, \mathbf{I})$ , где  $\mathbf{W} \in \mathbb{R}^{n \times d}$  матрица преобразования, выбранная Вами. Сравнить результат работы алгоритма из п. в) с обычным методом главных компонент для  $d = 2$  (10 баллов);

д) (автоматическое определение числа компонент) Считаем, что  $d = n$ . Введем априорное распределение на  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_n]$  вида

$$p(\mathbf{W} | \boldsymbol{\alpha}) = \prod_{j=1}^n \left( \sqrt{\frac{\alpha_j}{2\pi}} \right)^n \exp \left( -\frac{\alpha_j}{2} \mathbf{w}_j^\top \mathbf{w}_j \right),$$

где  $\mathbf{w}_j$  – столбцы матрицы  $\mathbf{W}$ .

Если  $\alpha_j \rightarrow \infty$ , то  $\mathbf{w}_j^\top \mathbf{w}_j \rightarrow 0$ , то есть происходит исключение соответствующей компоненты

из разложения  $\mathbf{x}_i = \mathbf{W}\mathbf{z}_i + \boldsymbol{\mu} + \boldsymbol{\varepsilon}$ , что соответствует сокращению числа главных компонент. С помощью вариационного EM-алгоритма решить задачу (50 баллов)

$$p(\mathbf{X}|\boldsymbol{\mu}, \sigma, \boldsymbol{\alpha}) \rightarrow \max_{\boldsymbol{\alpha}, \boldsymbol{\mu}, \sigma} .$$

**Подсказка:** в качестве скрытых переменных рассмотреть  $(\mathbf{Z}, \mathbf{W})$  и на E-шаге использовать вариационное приближение  $p(\mathbf{Z}, \mathbf{W}|\mathbf{X}, \boldsymbol{\mu}, \sigma, \boldsymbol{\alpha}) \approx q(\mathbf{Z})q(\mathbf{W})$ .

е) Для матрицы признаков из п. г) воспользоваться результатом пункта д) и проверить, происходит ли исключение восьми лишних главных компонент в ходе максимизации обоснованности (20 баллов).