Практическое задание 1 по курсу «Байесовский выбор моделей II»

Общая информация

- Время сдачи задания: 16е апреля, 23:59 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 практическому заданию #1 или решение практического задания #1;
- 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 (байесовская смесь моделей линейной регрессии). Пусть имеется K поставщиков одного товара, например, новой модели Xphone. У каждого поставщика с номером k есть базовая отпускная цена на этот товар p_k. Магазины в разных городах страны покупают этот товар у одного из поставщиков, причем вероятность выбора поставщика k есть π_k, $\boldsymbol{\pi} = [\pi_k,\ k = 1,\ \ldots,\ K]^\mathsf{T}$. Цена продажи поставщика отличается в зависимости от города, для которого магазин покупает товар и меняется с учетом следующих факторов:

- Уровня конкуренции;
- Покупательской способности;
- Уровня арендных ставок и т.д.

За каждую единицу проданного товара магазин получает от производителя фиксированную премию, но магазин на свое усмотрение может установить цену как ниже, так и выше, чем цена покупки у поставщика. Итоговая цена в магазине (которую мы наблюдаем) дается следующей моделью

$$P_i = p_{k_i} + \mathbf{v}_{k_i}^\mathsf{T}\mathbf{x}_i + \varepsilon_i = \mathbf{w}_{k_i}\mathbf{x}_i + \varepsilon_i,$$

где p_{k_i} – базовая цена выбранного поставщика, $\mathbf{x}_i$ – признаковое описание района, где размещен магазин, $\mathbf{w}_{k_i}$ – веса признаков в признаковом описании района, где размещен магазин, для выбранного поставщика, включая константный признак для учета базовой цены p_{k_i}, а ε_i – поправка к цене, устанавливаемая магазином. Считаем, что поправка для каждого магазина выбирается независимо от других магазинов, а также от того, какой поставщик товара был выбран, и в каком районе находится магазин.

Пусть имеется выборка $(\mathbf{X},\ \mathbf{p}) = (\mathbf{x}_i,\ P_i),\ i = 1,\ \ldots,\ m$ описаний разных магазинов, а также информация о цене на Xphone в них. Пусть K – оценка сверху на общее количество поставщиков, которое неизвестно (например, $K = 100$). В качестве априорного распределения на $\boldsymbol{\pi}$ введем распределение Дирихле $p(\boldsymbol{\pi}|\mu) = \mathrm{Dir}(\boldsymbol{\pi}|\mu\mathbf{e})$, где $\mu < 1$ для поощрения разреженности (например, $\mu = 10^{-6}$). На $\mathbf{w}_k$ введем априорное нормальное распределение $\mathbf{w}_k \sim \mathcal{N}(\mathbf{w}_k|\mathbf{0},\ \mathbf{A}_k^{-1}),\ k = \overline{1,\ K}$, где $\mathbf{A}_k$ – диагональная матрица. Шум (поправки к цене, устанавливаемые магазином) считаем нормальным, то есть $\varepsilon_i \sim \mathcal{N}(\varepsilon_i|0,\ \sigma_{k_i}^2)$ с дисперсией в зависимости от выбранного поставщика.

- Выписать совместное правдоподобие $p(\mathbf{p}, \mathbf{w}_1, \ldots, \mathbf{w}_K, \boldsymbol{\pi}|\mathbf{X}, \mathbf{A}, \boldsymbol{\sigma}^2, \mu)$ описанной модели в явном виде, где $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_K]$ (2 балла);

- Выписать апостериорное распределение $p(\boldsymbol{\pi}, \mathbf{w}_1, \ldots, \mathbf{w}_K|\mathbf{p}, \mathbf{X}, \mathbf{A}, \boldsymbol{\sigma}^2, \mu)$ с точностью до мультипликативной константы и качественно описать, почему не удается указать параметрический вид для этого распределения (2 балла);

- Ввести матрицу скрытых переменных $\mathbf{Z} = \|z_{ik}\|$, где $z_{ik} = 1$, если для i-го магазина выбран поставщик k, и выписать совместное правдоподобие модели со скрытой переменной $p(\mathbf{p}, \mathbf{w}_1, \ldots, \mathbf{w}_K, \boldsymbol{\pi}, \mathbf{Z}|\mathbf{X}, \mathbf{A}, \boldsymbol{\sigma}^2, \mu)$ (2 балла);

- Используя вариационное приближение $q(\boldsymbol{\pi}, \mathbf{w}_1, \ldots, \mathbf{w}_K, \mathbf{Z}) = q(\mathbf{Z})q(\boldsymbol{\pi})q(\mathbf{w}_1, \ldots, \mathbf{w}_K)$ для апостериорного распределения $p(\boldsymbol{\pi}, \mathbf{w}_1, \ldots, \mathbf{w}_K, \mathbf{Z}|\mathbf{p}, \mathbf{X}, \mathbf{A}, \boldsymbol{\sigma}^2, \mu)$ получить $q(\mathbf{Z}), q(\boldsymbol{\pi}), q(\mathbf{w}_1, \ldots$ в явном виде с формулами для параметров распределений (14 баллов);

- Используя принцип максимума обоснованности, провести оптимизацию гиперпараметров в смеси моделей путем решения задачи

$$p(\mathbf{p}|\mathbf{X}, \mathbf{A}, \boldsymbol{\sigma}^2, \mu) \to \max_{\mathbf{A}, \boldsymbol{\sigma}^2}$$

  с помощью вариационного EM-алгоритма, считая μ известным и фиксированым (15 баллов). Описать при каких условиях признак j исключается из k-й модели в смеси, то есть происходит отбор признаков (2 балла);

- Реализовать вариационный EM-алгоритм и подсчет нижней оценки обоснованности (ELBO), максимизируемой в EM-алгоритме (3 балла, см. ноутбук с лекции 3). Сгенерировать синтетическую выборку из смеси 3 моделей (2 балла, модифицировать генератор из ноутбука) и для полученной выборки запустить вариационный EM-алгоритм и описать результаты (3 балла);

- В эксперименте из предыдущего пункта проследить за изменением ELBO на кажой подитерации EM-алгоритма и убедиться, что значение ELBO растет (2 балла). Попробовать «испортить» формулы каждого из шагов и проследить, появляются ли итерации, на которых ELBO снижается. Если таковых нет на данной выборке, сгенерировать несколько новых и исследовать долю тех, где неверные формулы приводят к снижению ELBO (5 баллов);

- Сгенерировать тестовую выборку из той же модели (1 балл). Воспользоваться полученными апостериорными распределениями параметров смеси для прогноза на тестовой выборке и описать полученный результат (3 балла).

- Предположим, что про часть магазинов стало известно, к какому поставщику они относятся. Выписать правдоподобие со скрытыми переменными для такой модели (1 балл). Выписать поправки к вариационному EM-алгоритму на этот случай (3 балла) и поправки к ELBO (2 балла);

- Выбрав случайно половину объектов (магазинов) и считая, что для них известен настоящий поставщик k_i, применить вариационный EM-алгоритм к выборке с наполовину известными скрытыми переменными (5 балла);

- Для той же тестовой выборки с неизвестными поставщиками сравнить качество прогноза с полученным ранее (4 балла). Уменьшить число объектов в обучающей выборки и повторить сравнение (4 балла).

- Вернуться к исходной синтетической выборке и изучить зависимость результата от начального приближения - использовать 100 разных начальных приближений (постарайтесь их сделать заметно разными). Сходится ли алгоритм в ту же точку? Как сильно отличается значение ELBO в локальных оптимумах? (7 баллов) На основании полученных результатов решить, нужен ли мультистарт и если да, как выбрать среди локальных оптимумов решение (3 балла);

- Измерить время работы вариационного EM-алгоритма на синтетической выборке (1 балл). Предложить улучшение критерия останова для сокращения времени работы (3 балла) и измерить ускорение (2 балла). Предложить способ раннего останова запусков из неперспективного начального приближения (2 балла) и оценить качество его применения (2 балла).

- **Тест на внешней выборке.** Выборка включает в себя обучающую совокупность $(\mathbf{X}_{\text{train}}, \mathbf{p}_{\text{train}})$, а также признаковое описание тестовой совокупности $\mathbf{X}_{\text{test}}$. Требуется построить прогноз $\hat{\mathbf{p}}_{\text{test}}$, а также вектор неуверенности в прогнозе $\mathbf{u}_{\text{test}}$ оптимальные в следующем смысле
$$-\tfrac{1}{2}\sum_{i=1}^{m_2} \log u_i^2 - \tfrac{1}{2u_i^2}(P_i - \hat{P}_i)^2 \to \max,$$
где суммирование производится по объектам тестовой выборки. В выборке также имеются данные о владельцах магазинов - подумайте как их использовать (7 баллов + 30 баллов наилучшему прогнозу, 20 баллов второму по качеству).