

# Вопросы

Качанов Владимир, Стрелкова Евгения

25 ноября 2018 г.

**1** Что такое логистическая регрессия? Какие проблемы она имеет? Что происходит с оценкой максимума правдоподобия для вектора параметров лог. регрессии, если выборка идеально разделима?

**Ответ:** Метод построения линейного классификатора, позволяющий оценивать апостериорные вероятности принадлежности объектов классам. В случае идеально разделимой выборки существует множество векторов идеально разделяющих выборку, сигмоидная функция не определена. Максимум правдоподобия не определен, так как  $w_{ML}$  имел бы бесконечную норму.

**2** Что такое регуляризация? Для чего она нужна? Чему она соответствует в байесовской статистике?

**Ответ:** Метод добавления некоторой дополнительной информации к условию с целью решить проблему переобучения, имеет вид штрафа за сложность модели. Введение регуляризации соответствует введению априорного распределения:  $l_1$ -регуляризация соответствует априорному распределению Лапласа, а  $l_2$ -регуляризация - нормальному распределению.

**3** Что такое априорное распределение (prior)? Для чего оно нужно? Как выбрать prior?

**Ответ:** Априорное распределение - это распределение вероятностей до того как мы узнали выборку. Нужно, например, для определения апостериорного распределения. Либо в случае мультимоделей, если мы очень сильно (не) верим в какую-то из моделей. Prior можно выбрать основываясь на предпочтениях, знаниях вопроса или опыте.

**4** Что такое апостериорное распределение (posterior)? Как оно связано с prior и правдоподобием? При каком условии (сопряженность (conjugate prior)) posterior лежит в том же классе распределений, что и prior? (например, нормальное распределение -> нормальное, гамма -> гамма)

**Ответ:**  $p(y|X, w) = p(y|X, w)p(w|\alpha)$  - правдоподобие, где  $p(w|\alpha)$  - априорное распределение  
 $p(w|X, y, \alpha) = \frac{p(y|X, w)p(w|\alpha)}{p(y|X, \alpha)}$  - апостериорное распределение.

Если  $p(x)$  представимо в виде  $p(x|\Theta) = \frac{h(x)}{Z(\Theta)} \exp(\Theta^T u(x))$ , то  $p(\Theta) = \frac{H(\alpha, v)}{Z(\Theta)^\alpha} \exp(\Theta^T v)$  будет давать апостериорное распределение того же типа (сопряженное). В общем случае, нам необходимо подобрать априорное распределение так, чтобы при добавлении его логарифма к квадратичной форме, получилось то же, что и было в априорном распределении.

**5** Что такое обоснованность? Какой ее физический смысл?

**Ответ:**  $p_i(y|X) = \int p_i(y|X, w)p_i(w)dw$  - обоснованность модели. Обоснованность помогает определить сложность модели, например, более простую в случае малого объема выборки, а значит отсутствия статистической значимости выбора более сложной, либо сложную, если выборка большая, и любые особенности выборки являются статистически значимыми.

**6** Как связана обоснованность и апостериорная вероятность модели, когда выборка получена из одной из списка моделей?

**Ответ:** Апостериорное распределение модели:  $p(M_i|X_{train}, y_{train}) = \frac{p(y_{train}, M_i|X_{train})}{p(y_{train}|X_{train})} \sim p(y_{train}, M_i|X_{train}) = \int p(y_{train}, M_i, w|X_{train})dw = p(M_i)p_i(y_{train}|X_{train})$ , где  $p(M_i)$  - априорное распределение моделей, а  $p_i(y_{train}|X_{train})$  - обоснованность модели. Чем больше обоснованность модели, тем более вероятна апостериори модель. Если нет никаких априорных предпочтений по поводу модели, то наиболее вероятна та модель, которая обладает наибольшей обоснованностью.

**7** Что такое принцип максимума обоснованности? При каком условии он дает тот же результат, что и "честное" апостериорное распределение на прогноз с учетом наличия многих моделей?

**Ответ:** Принцип максимума обоснованности это принцип выбора модели, когда мы берем только модель име-

ющую наибольшую обоснованность, а не взвешенную сумму всех. Он дает тот же результат что и "честный" подсчет в случае, когда одна модель значительно более обоснованна чем все остальные.

**8** Есть ли сопряженность нормального априорного распределения с правдоподобием логистической регрессии? Существует ли вообще сопряженный prior для логистической регрессии?

**Ответ:** Для логистической регрессии сопряженный prior выбрать нельзя, т.к.  $w$  стоит в экспоненте в дроби. Чтобы задать такое априорное распределение, потребуется число параметров, равное числу объектов выборки. Соответственно, для нормального априорного распределения также не сопряжено.

**9** От чего зависит обоснованность модели логистической регрессии, если prior на вектор параметров  $N(0, A^{-1})$ , где  $A$  - диагональная матрица? Что происходит, если  $\alpha_i \rightarrow \infty (A = (\vec{\alpha}))$ ?

**Ответ:** Если  $\alpha_i \rightarrow \infty$ , то соответствующий параметр уходит в 0, значит он нам не нужен.