

Кластеризация и ЕМ-алгоритм

НИС Машинное обучение и приложения, 2016/2017

1. Чем отличаются задачи обучения с учителем и без учителя (supervised/unsupervised learning)? В чём сложность задачи обучения без учителя?
2. В чём разница между soft и hard clustering? Почему soft clustering может быть предпочтительнее?
3. Что такое апостериорная вероятность?
4. В терминах параметрической модели с латентными переменными, что является наблюдаемой, а что скрытой переменной для смеси гауссиан? Как выглядят априорное и апостериорное распределения на скрытую переменную?
5. Предположим, в параметрической модели есть наблюдаемые переменные X и скрытые переменные Z , а правдоподобие $P(X|\theta)$ нельзя максимизировать напрямую. В чём основная идея ЕМ-алгоритма (как он решает последнюю проблему)?
6. Что такое KL-дивергенция?