**Контрольные вопросы**

1. **Почему байесовский алгоритм называют наивным?**

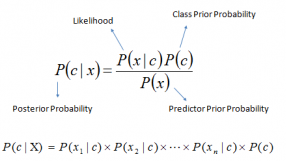
В машинном обучении наивный байесовский алгоритм – это алгоритм классификации, основанный на теореме Байеса. Он называется наивным, потому что в основе этого алгоритма лежат наивные предположения. Вот некоторые из преимуществ этого алгоритма:

Это очень простой алгоритм для задач классификации по сравнению с другими алгоритмами классификации.

1. **Чем статистический подход к классификации отличается от детерминистского?**

Cтатистическом подходе, в отличие от детерминистских, для которых характерно однозначное отнесение образа к одному из классов, полагается, что объект может принадлежать любому из классов, но с некоторой вероятностью

1. **Приведите теорему Байеса, используемую для классификации.**

****

1. **Что такое правдоподобие?**

*P*(*x*|*c*) – плотность распределения вероятностей образов класса *с* или правдоподобие того, что из класса *с* будет выбран образ *x*.

1. **Каким в самом простом случае при построении байесовских классификаторов является представление плотностей распределений условных вероятностей образов классов?**

В случае независимости признаков матрица ковариации – диагональная, по главной диагонали стоят дисперсии компонент вектора признаков, *P*(*x*|*с*) является произведением n одномерных плотностей. Предположение о независимости существенно упрощает задачу, так как оценить n одномерных плотностей гораздо проще, чем одну *n*-мерную плотность.

1. **Перечислите преимущества и недостатки НБА.**

***Преимущества:***

* классификация, в том числе многоклассовая, выполняется легко и быстро;
* когда допущение о независимости выполняется, НБА превосходит другие алгоритмы, такие как логистическая регрессия (logistic regression), и при этом требует меньший объем обучающих данных;
* НБА лучше работает с категорийными признаками, чем с непрерывными. Для непрерывных признаков предполагается нормальное распределение, что является достаточно сильным допущением.

***Недостатки:***

Если в тестовом наборе данных присутствует некоторое значение категорийного признака, которое не встречалось в обучающем наборе данных, тогда модель присвоит нулевую вероятность этому значению и не сможет сделать прогноз. Это явление известно под названием «нулевая частота» (zero frequency). Данную проблему можно решить с помощью сглаживания. Одним из самых простых методов является сглаживание по Лапласу (Laplace smoothing).

Хотя НБА является хорошим классификатором, значения спрогнозированных вероятностей не всегда являются достаточно точными.

Еще одним ограничением НБА является допущение о независимости признаков. В реальности наборы полностью независимых признаков встречаются крайне редко.