В практическом аспекте, когда мы не можем в полной мере оперировать понятием события, опеределяемым классической теорией вероятности, приходит на помощь байесовский подход, основанный на понятии гипотезы. Наше событие становится гипотезой, а вероятность данного события – степенью доверия к этой гипотезе. Разумеется, эту вероятность возможно определить по формуле Байеса как отношение априорной вероятности события, умноженной на степень доверия к ним, к полной вероятности события.

Этот же подход имеет место и в машинном обучении. Имеем же здесь, конечно, не саму формулу Байеса, но то же по виду. Итак, рассмотрим задачу многоклассовой классификации на n непересекающихся классов. Чем же обычная сигмоида в данном случае хуже. Самое главное отличие, что softmax – нормированная функция, то есть сумма вероятностей принадлежностей к кадому классу по всему пространству гипотез в точности равна 1, чего не наблюдается в случае сигмоидальной функции. Можно привести пример, но я не буду. Во-вторых, есть следующее замечание: большему значению будет соответствовать большее значение функции и большая вероятность, чего не наблюдается в сигмоиде, потому что у неё маленький рабочий отрезок.

**Упражнение**

Показать, что в случае бинарной классификации сигмоида и софтмах – это одно и то же.

Обязательно построй графики функций!

Источник: https://glassboxmedicine.com/2019/05/26/classification-sigmoid-vs-softmax/