

Обучение с учетом издержек классификации (Cost-sensitive learning)

Методика обучения модели бинарной классификации в условиях классификации с учетом издержек с целью минимизации ошибок I или II рода (ложно-положительных или ложно-отрицательных), в зависимости от того, какие из них ведут к большим издержкам, связанным с принятием неверных управленческих решений.

Наиболее простым способом является изменение доли положительных или отрицательных примеров в обучающем множестве. Например, если большие издержки связаны с ложно-положительными ошибками, то можно искусственно увеличить долю отрицательных примеров. Тогда модель, стремящаяся в процессе обучения минимизировать число ошибок, «научится» лучше распознавать отрицательные примеры и будет допускать на них меньше ошибок (ложно-положительных).

Еще одним способом снижения вероятности ошибок, связанных с большими издержками, является подбор дискриминационного порога бинарных классификаторов.