

Принцип максимальной энтропии (Maximum entropy method)

Принцип максимальной энтропии — это понятие теории вероятности, представляющее собой утверждение, что распределение вероятностей, которое наилучшим образом отражает текущее состояние данных — это распределение с наибольшей информационной энтропией.

В анализе данных этот принцип используется для построения статистических классификаторов. Как известно, энтропия — это мера неопределенности некоторого вероятностного распределения:

$$H(x) = - \sum p(x) \log p(x)$$

С практической точки зрения, чем выше энтропия как мера неопределенности, тем сложнее делать предсказания класса. Поэтому можно предположить, что для улучшения качества классификации энтропию множества нужно уменьшать. В пределе мы получим единственный класс, предсказание которого будет полностью достоверным.

Противоположной будет ситуация, когда классы равновероятны и неопределенность классификации максимальна, как и энтропия.

Классификаторы на основе критерия максимума энтропии похожи на логистическую регрессию. Только минимизируется в этом случае не логарифмическое правдоподобие, а энтропия.

Принцип был впервые изложен Э. Т. Джейнсом в 1957 г.