Три математических кита ИИ. Теория вероятностей и математическая статистика

Сириус, 03.07.2025

Спикер: Гасников А.В, ректор университета Иннополис

Оглавление

Ключевые термины:

искусственный интеллект, формула Байеса, центральная предельная теорема, максимальное правдоподобие, случайные величины, нормальное распределение, машинное обучение

Введение

Вероятностные модели играют ключевую роль в современном искусственном интеллекте. Они позволяют описывать неопределённость в данных, строить адаптивные системы и решать задачи обучения в условиях шума. В этой лекции рассматриваются фундаментальные понятия: формула Байеса, центральная предельная теорема, нормальное распределение и принцип максимального правдоподобия, а также их связь с машинным обучением.

Вероятностные модели

Вероятностная модель — математическое описание, где неопределённость выражается через вероятности. В задачах ИИ такие модели используются при предсказании, принятии решений и обучении агентов. Они включают в себя случайные величины, функции плотности и методы оценки параметров.

Формула Байеса

Формула Байеса позволяет обновлять апостериорные вероятности на основе новых данных:

$$\boxed{P(A|B) = (P(B|A) * P(A)) / P(B)}$$

Применение: задача многоруких бандитов. Пусть каждый рычаг имеет вероятность успеха, которая априорно неизвестна. После каждого испытания вероятность обновляется через бета-распределение:

$$\boxed{P(\text{успех}_k) = (s_k + 1) / (n_k + 2)}$$

где s_k — число успехов, n_k — число испытаний.

Случайные величины, математическое ожидание и дисперсия

Случайная величина X принимает значения в зависимости от результата случайного эксперимента.
Математическое ожидание:

$E[X] = \Sigma x_i * P(X = x_i)$ (дискретная)
$E[X] = \int x * p(x)\, dx$ (непрерывная)
Дисперсия измеряет разброс значений:
$Var(X) = E[(X - E[X])^2]$
Свойства:
- $E[X + Y] = E[X] + E[Y]$
- $Var(X + Y) = Var(X) + Var(Y)$, если X и Y независимы

Центральная предельная теорема

ЦПТ утверждает, что сумма n независимых одинаково распределённых случайных величин с ожиданием μ и дисперсией σ^2 имеет приближённо нормальное распределение при $n \to \infty$:
$(\Sigma X_i - n\mu) / (\sigma\sqrt{n}) \to N(0, 1)$

Это объясняет, почему нормальное распределение так часто возникает в статистике.

Нормальное распределение

Формула плотности:

$p(x) = (1 / (\sigma\sqrt{2\pi})) * e\hat{}(-(x - \mu)^2 / (2\sigma^2))$

Нормальное распределение устойчиво к сложению: сумма нормальных величин — тоже нормальная. Часто используется как модель шума в данных. Оценка методом наименьших квадратов основана на этом предположении.

Максимальное правдоподобие

$L(\theta) = \Pi\, p(x_i \mid \theta)$, $i=1..n$
Оценка $\hat{\theta}$ максимизирует $L(\theta)$. Часто используют логарифм:
$\log L(\theta) = \Sigma \log p(x_i \mid \theta)$

Принцип максимального правдоподобия: оценка параметров модели так, чтобы вероятность наблюдаемых данных была максимальной:

Пример: вероятность успеха монеты = число успехов / общее число бросков.

Статистика и машинное обучение

Статистика даёт математический аппарат для анализа моделей и оценки их параметров. Машинное обучение ориентировано на минимизацию ошибок:
- Нейросети — модели с множеством параметров
- Регуляризация (например, L2) предотвращает переобучение
- Качество модели оценивается на тестовой выборке

Примеры применения

Примеры из реальной жизни:
- Многорукие бандиты (обучение с подкреплением)
- Оценка числа заболевших COVID через выборку
- Рекомендательные системы (например, фильмы)
- Оценка плотности методом гистограмм
- Прогнозирование голосования

Роль больших данных

Больше данных — выше точность:
- Точность оценки растёт как $1/\sqrt{n}$
- Данные должны быть репрезентативны
- Деление на train/val/test снижает переобучение

Заключение

Вероятностные методы и статистика — фундамент ИИ. Они позволяют строить интерпретируемые и надёжные алгоритмы. Понимание формул, распределений и методов оценки — ключ к успешной работе с данными и моделями.