Конспект лекции: *«Три математических кита ИИ: Теория вероятностей и математическая статистика»*

Гасников Александр Владимирович
Ректор Университета Иннополис, профессор МФТИ,
доктор физико-математических наук

4 июля 2025 г.

Содержание

1 Три фундаментальных раздела математики для ИИ

1. **Теория вероятностей и статистика**: формула Байеса, центральная предельная теорема (ЦПТ), оценка максимального правдоподобия (ОМП)

2. **Оптимизация**: принципы Ферма, Лагранжа, градиентный спуск

3. **Линейная алгебра**: теорема Рисса, сингулярное разложение (SVD), метод главных компонент (PCA)

2 Ключевые концепции вероятности и статистики

2.1 Случайные величины и распределения

- **Случайная величина (СВ)** — функция, отображающая исходы эксперимента в числа.
- **Плотность распределения** $p(x)$:
  - Равномерное распределение на $[0,1]$: $p(x) = 1$
  - Биномиальный эксперимент (задача многоруких бандитов):

$$p_k(x) = \frac{(l_k + w_k + 1)!}{l_k!\, w_k!} x^{w_k}(1-x)^{l_k}, \quad \mathbb{E}[P_k] = \frac{w_k + 1}{l_k + w_k + 2}$$

2.2 Математическое ожидание и дисперсия

- **Математическое ожидание**:

$$\mathbb{E}[\xi] = \sum \xi_i p_i, \quad \mathbb{E}\left[\sum \xi_k\right] = \sum \mathbb{E}[\xi_k]$$

- **Дисперсия**:

$$\mathrm{Var}(\xi) = \mathbb{E}[(\xi - \mathbb{E}[\xi])^2], \quad \mathrm{Var}\left(\sum \xi_k\right) = \sum \mathrm{Var}(\xi_k) \quad \text{(для независимых СВ)}$$

2.3 Центральная предельная теорема (ЦПТ)

- **Формулировка**:

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad \frac{1}{n} \sum_{k=1}^{n} \xi_k = m + \frac{\sigma}{\sqrt{n}}\eta, \quad \eta \sim \mathcal{N}(0,1)$$

- **Доверительный интервал**:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum \xi_k - h\right| \leq \frac{\sigma}{\sqrt{n}}\sqrt{\ln(1/\delta)}\right) \geq 1 - \delta$$

3 Статистические методы и их приложения

3.1 Принцип максимума правдоподобия (ОМП)

- **Общая формула**:

$$\hat{\theta} = \arg\max_{\theta} \prod_{k=1}^{n} p(\xi_k | \theta)$$

- **Пример: монетка**:

$$L = x^S (1-x)^{n-S}, \quad \hat{x} = \frac{S}{n}$$

3.2 Метод наименьших квадратов (МНК)

- Решение задачи линейной регрессии:

$$\hat{h} = \arg\min_{h} \sum_{k=1}^{n} (v_k - h \cdot r_k)^2$$

3.3 Байесовский подход

- **Базовая формула**:
$$\hat{x} = \arg\max_x \prod p(\xi_k|x) \cdot \Pi(x)$$

- **Нормальное распределение**:
$$\hat{x} = \arg\min \left[\frac{1}{2\sigma_\xi^2} \sum (\xi_k - x)^2 + \frac{x^2}{2\sigma_x^2} \right]$$

- **LASSO-регуляризация**:
$$\hat{x} = \arg\min \left[\frac{1}{2\sigma_\xi^2} \sum (\xi_k - x)^2 + \lambda |x| \right]$$

4 Машинное обучение и математическая статистика

4.1 Обучение нейронных сетей

- **Статистическая модель**:
$$y = f(a, x) + \xi, \quad \xi \sim \mathcal{N}(0, \sigma^2), \quad \min_x \mathbb{E}_{(a,y)} \left[(y - f(a, x))^2 \right]$$

- **ML-подход (эмпирический риск)**:
$$\min_x \frac{1}{n} \sum_{k=1}^{n} (y_k - f(a_k, x))^2$$

4.2 Проблема переобучения

- **Сравнение моделей**:
$$_{f_1} = \frac{1}{|\Delta_2|} \sum_{k \in \Delta_2} (y_k - f_1(a_k, x_1))^2$$
$$_{f_2} = \frac{1}{|\Delta_2|} \sum_{k \in \Delta_2} (y_k - f_2(a_k, x_2))^2$$

5 Философские и практические выводы

- Точность оценки: $\sim \frac{1}{\sqrt{n}}$
- Стохастическая оптимизация:
$$\min_x \mathbb{E}[\mathcal{L}(y, f(a, x))]$$

- Рост числа параметров требует роста объёма данных:
$$\uparrow \text{Параметры} \quad \Rightarrow \quad \uparrow \text{Данные}$$

Заключение

- Центральная предельная теорема объясняет возникновение нормальных распределений.
- Оценка максимального правдоподобия обладает оптимальными свойствами.
- Разделение данных на обучающую и тестовую выборки критически важно.
- Регуляризация помогает избежать переобучения.
- Математическая статистика лежит в основе корректного машинного обучения.