Байесовские Методы. Лекция 2

БИВТ-23-9/10-ИСАД

15 сентября 2025 г.

Как рождаются научные объяснения

- **Наблюдения** $\rightarrow$ формулировка **гипотез**.

Как рождаются научные объяснения

- **Наблюдения** $\to$ формулировка **гипотез**.
- **Предсказуемость**: хорошая гипотеза объясняет имеющиеся данные и делает проверяемые предсказания.

Как рождаются научные объяснения

- **Наблюдения** $\rightarrow$ формулировка **гипотез**.
- **Предсказуемость**: хорошая гипотеза объясняет имеющиеся данные и делает проверяемые предсказания.
- **Экономия допущений**: из двух сопоставимых объяснений предпочитаем *менее нагруженное* (меньше сущностей/свобод).

Как рождаются научные объяснения

- **Наблюдения** $\to$ формулировка **гипотез**.
- **Предсказуемость**: хорошая гипотеза объясняет имеющиеся данные и делает проверяемые предсказания.
- **Экономия допущений**: из двух сопоставимых объяснений предпочитаем *менее нагруженное* (меньше сущностей/свобод).
- **Баланс** объяснительной силы и простоты — ключ к обобщающей способности.

Бритва Оккама: интуиция

Принцип

«Не умножай сущности без необходимости»: среди гипотез, одинаково хорошо объясняющих данные, выбирай **более простую**.

- **Не про наивность**, а про *минимум избыточных допущений*.

Бритва Оккама: интуиция

Принцип

«Не умножай сущности без необходимости»: среди гипотез, одинаково хорошо объясняющих данные, выбирай **более простую**.

- **Не про наивность**, а про *минимум избыточных допущений*.
- **Сложность** оправдана, только если она улучшает объяснение/предсказание.

Бритва Оккама: интуиция

Принцип

«Не умножай сущности без необходимости»: среди гипотез, одинаково хорошо объясняющих данные, выбирай **более простую**.

- **Не про наивность**, а про *минимум избыточных допущений*.
- **Сложность** оправдана, только если она улучшает объяснение/предсказание.
- Пример: модель монетки «честная» vs «каждый запуск подстраивается» — второе объясняет всё, но *обобщает плохо*.

Байесовская формализация выбора моделей

Теорема Байеса для моделей

$$P(M \mid D) = \frac{P(D \mid M)\, P(M)}{P(D)}$$

- $P(M)$ — априорная вероятность модели.

Байесовская формализация выбора моделей

Теорема Байеса для моделей

$$P(M \mid D) = \frac{P(D \mid M)\, P(M)}{P(D)}$$

- $P(M)$ — априорная вероятность модели.
- $P(D \mid M)$ — **model evidence** (маргинальное правдоподобие):

$$P(D \mid M) = \int P(D \mid \theta, M)\, P(\theta \mid M)\, d\theta$$

Байесовская формализация выбора моделей

Теорема Байеса для моделей

$$P(M \mid D) = \frac{P(D \mid M)\, P(M)}{P(D)}$$

- $P(M)$ — априорная вероятность модели.
- $P(D \mid M)$ — **model evidence** (маргинальное правдоподобие):

$$P(D \mid M) = \int P(D \mid \theta, M)\, P(\theta \mid M)\, d\theta$$

- $P(D)$ — нормировочный множитель (одинаковый для всех моделей).

Байесовская формализация выбора моделей

Теорема Байеса для моделей

$$P(M \mid D) = \frac{P(D \mid M)\,P(M)}{P(D)}$$

- $P(M)$ — априорная вероятность модели.
- $P(D \mid M)$ — **model evidence** (маргинальное правдоподобие):

$$P(D \mid M) = \int P(D \mid \theta, M)\,P(\theta \mid M)\,d\theta$$

- $P(D)$ — нормировочный множитель (одинаковый для всех моделей).
- Выбираем модель с наибольшим $P(M \mid D)$.

Байесовская формализация выбора моделей

Теорема Байеса для моделей

$$P(M \mid D) = \frac{P(D \mid M)\, P(M)}{P(D)}$$

- $P(M)$ — априорная вероятность модели.
- $P(D \mid M)$ — **model evidence** (маргинальное правдоподобие):

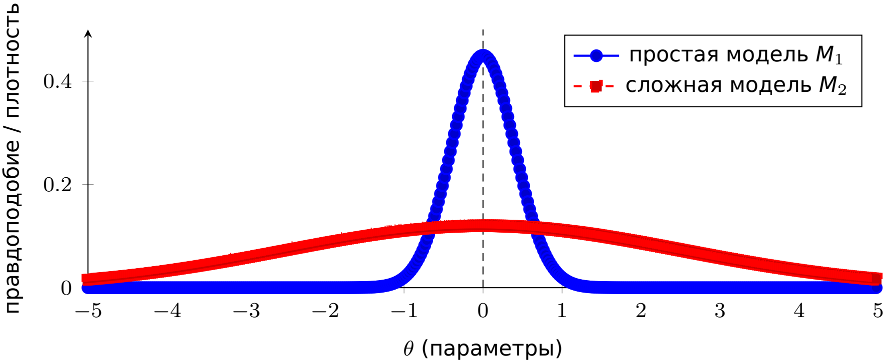
$$P(D \mid M) = \int P(D \mid \theta, M)\, P(\theta \mid M)\, d\theta$$

- $P(D)$ — нормировочный множитель (одинаковый для всех моделей).
- Выбираем модель с наибольшим $P(M \mid D)$.

Интуиция (бритва Оккама)

Evidence усредняет правдоподобие по всем параметрам. Сложные модели имеют больше «пустого объёма параметров» → вероятность размазывается → штраф за сложность.

Пример: почему сложная модель получает меньший evidence



- Обе модели могут объяснить данные около $\theta = 0$.
- Но интеграл $P(D \mid M)$ для M_2 меньше — вероятность «размазана» по большому пространству параметров.
- Байесовский выбор отдаст предпочтение M_1 — **формализация бритвы Оккама**.

Дискриминативные vs Генеративные модели

- **Дискриминативные модели**: напрямую учат связь между входом и выходом.

  $$P(y \mid x)$$

  Цель: хорошо предсказывать метки y для новых x.

Дискриминативные vs Генеративные модели

- **Дискриминативные модели**: напрямую учат связь между входом и выходом.
$$P(y \mid x)$$
Цель: хорошо предсказывать метки y для новых x.
- **Генеративные модели**: учат полное распределение данных.
$$P(x, y) \quad \text{или} \quad P(x \mid y)P(y)$$
Цель: моделировать процесс генерации данных (включая возможность сэмплирования новых x).

Что происходит при обучении

Дискриминативная модель

- Дано: обучающая выборка (x, y).
- Неизвестно: параметры θ.
- Оптимизация: максимизация $P(y \mid x; \theta)$.
- Валидация: проверка качества предсказаний y.

Генеративная модель

- Дано: обучающая выборка (x, y) или только x.
- Неизвестно: параметры распределений.
- Оптимизация: максимизация $P(x, y; \theta)$ или $P(x \mid \theta)$.
- Валидация: проверка, насколько модель воспроизводит распределение данных.

Как выбирать модель?

- В классическом ML часто выбираем модель по:

Как выбирать модель?

- В классическом ML часто выбираем модель по:
    - качеству на тренировке,

Как выбирать модель?

- В классическом ML часто выбираем модель по:
  - качеству на тренировке,
  - качеству на валидации,

Как выбирать модель?

- В классическом ML часто выбираем модель по:
  - качеству на тренировке,
  - качеству на валидации,
- Но эти подходы — эвристики, не всегда строго обоснованы.

Как выбирать модель?

- В классическом ML часто выбираем модель по:
  - качеству на тренировке,
  - качеству на валидации,
- Но эти подходы — эвристики, не всегда строго обоснованы.
- В байесовском подходе естественный критерий выбора модели — **model evidence**.

$$P(M \mid D) \propto P(D \mid M) \cdot P(M)$$

Как выбирать модель?

- В классическом ML часто выбираем модель по:
  - качеству на тренировке,
  - качеству на валидации,
- Но эти подходы — эвристики, не всегда строго обоснованы.
- В байесовском подходе естественный критерий выбора модели — **model evidence**.

$$P(M \mid D) \propto P(D \mid M) \cdot P(M)$$

- Выбираем модель с максимальным апостериорным $P(M \mid D)$.

Maximum Evidence Method (Type-II MLE)

Определение

Для модели M с параметрами θ:

$$P(D \mid M) = \int P(D \mid \theta, M)\, P(\theta \mid M)\, d\theta$$

— **marginal likelihood / evidence**.

- Выбор модели:

$$\hat{M} = \arg\max_{M} P(D \mid M)$$

Maximum Evidence Method (Type-II MLE)

Определение

Для модели M с параметрами θ:

$$P(D \mid M) = \int P(D \mid \theta, M)\, P(\theta \mid M)\, d\theta$$

— **marginal likelihood / evidence**.

- Выбор модели:
$$\hat{M} = \arg\max_{M} P(D \mid M)$$
- Этот метод называют также **Type-II Maximum Likelihood Estimation**.

Maximum Evidence Method (Type-II MLE)

Определение

Для модели M с параметрами θ:

$$P(D \mid M) = \int P(D \mid \theta, M)\, P(\theta \mid M)\, d\theta$$

— **marginal likelihood / evidence**.

- Выбор модели:

$$\hat{M} = \arg\max_{M} P(D \mid M)$$

- Этот метод называют также **Type-II Maximum Likelihood Estimation**.
- Идея: выбираем гиперпараметры или модель в целом так, чтобы данные были наиболее вероятны в среднем по всем параметрам.

Maximum Evidence Method (Type-II MLE)

Определение

Для модели M с параметрами θ:

$$P(D \mid M) = \int P(D \mid \theta, M)\, P(\theta \mid M)\, d\theta$$

— **marginal likelihood / evidence**.

- Выбор модели:

$$\hat{M} = \arg\max_{M} P(D \mid M)$$

- Этот метод называют также **Type-II Maximum Likelihood Estimation**.
- Идея: выбираем гиперпараметры или модель в целом так, чтобы данные были наиболее вероятны в среднем по всем параметрам.
- В отличие от MLE (по θ), мы оптимизируем *на уровне моделей*.

Evidence для разных моделей монетки

Наблюдения: в n бросках выпало k «орлов». Правдоподобие:

$$P(D \mid \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

Модель M_1 (фиксированная честная монетка, $\theta = 0.5$):

$$P(D \mid M_1) = \binom{n}{k} 0.5^{\,n}.$$

Модель $M(\alpha, \beta)$ (неизвестная θ с бета-априором):

$$P(\theta \mid M) = \mathrm{Beta}(\alpha, \beta), \qquad P(D \mid M) = \int P(D \mid \theta)\, P(\theta \mid M)\, d\theta = \binom{n}{k} \frac{B(k+\alpha,\ n-k+\beta)}{B(\alpha, \beta)}.$$

Численный пример: $n = 20,\ k = 14$

$$P(D \mid M_1) = \binom{20}{14} 0.5^{20} \ \approx\ 0.0369644,$$

$$P(D \mid M_2) = \binom{20}{14} \frac{B(14+1,\ 6+1)}{B(1,1)} \ =\ \binom{20}{14} B(15,7) \ \approx\ 0.0476190,$$

$$P(D \mid M_3) = \binom{20}{14} \frac{B(14+2,\ 6+2)}{B(2,2)} \ =\ \binom{20}{14} \frac{B(16,8)}{1/6} \ \approx\ 0.0592885.$$

Байес-факторы (сравнение моделей):

$$\mathrm{BF}_{2,1} = \frac{P(D \mid M_2)}{P(D \mid M_1)} \approx 1.288, \qquad \mathrm{BF}_{3,1} \approx 1.604, \qquad \mathrm{BF}_{3,2} \approx 1.245.$$

- При этих данных evidence предпочитает бета-модели перед фиксированной «честной» монеткой.
- Более информативный априор $\mathrm{Beta}(2,2)$ даёт ещё большее evidence, чем $\mathrm{Beta}(1,1)$.

Пример: три гауссианские модели

- Рассмотрим пространство:
  - ось X — параметры θ,
  - ось Y — данные D.
- Есть три модели M_1, M_2, M_3.
- Каждая модель задаёт априор $P(\theta \mid M_i)$ (на X).
- Каждая модель через правдоподобие $P(D \mid \theta, M_i)$ даёт предсказания на оси Y.
- Пусть наблюдаемое значение данных y^* отмечено на оси Y.

Пример: три гауссианские модели