Семинар по байесовским методам

Д.А. Кропотов К.А. Струминский

Школа по байесовским методам в глубинном обучении

Задача 1

О точности медицинского теста

Дано

Тест имеет высокую точность:

- вероятность позитивного ответа при наличии заболевания составляет 99%
- вероятность отрицательного ответа при отсутствии заболевания также составляет 99%

В то же время выявленное заболевание является достаточно редким и встречается только у одного человека на 10000

Найти

Вероятность того, что у обследуемого человека действительно есть выявленное заболевание

Задача 1

Решение

Пусть $x = 1$ если человек болен, $x = 0$ в противном случае, $y = 1$ если тест дал положительный результат, $y = 0$ в противном случае.
Вычислим $p(x = 1|y = 1)$:

$$p(x = 1|y = 1) = \frac{p(x = 1, y = 1)}{p(y = 1)} = \frac{p(y = 1|x = 1)p(x = 1)}{\sum_j p(y = 1|x = j)p(x = j)} =$$

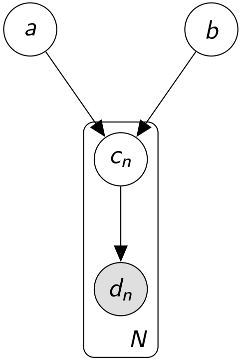
$$= \frac{0.99 \cdot 10^{-4}}{0.99 \cdot 10^{-4} + 10^{-2} \cdot 0.9999} = \frac{0.99}{0.99 + 99.99} \approx \frac{1}{100}$$

Задача 2

Вероятностная модель семинара на летней школе

Дано

$$p(a, b, c_1, \dots, c_N, d_1, \dots, d_N) =$$
$$= p(a)p(b)\prod_{n=1}^{N} p(d_n|c_n)p(c_n|a, b),$$
$$d_n|c_n \sim \mathrm{Bin}(c_n, \delta),$$
$$c_n|a, b \sim \mathrm{Bin}(a, \gamma_a) + \mathrm{Bin}(b, \gamma_b),$$
$$a \sim \mathrm{Bin}(N_a, \alpha),$$
$$b \sim \mathrm{Bin}(N_b, \beta).$$



Найти

Выражения для вычисления $p(a|d_1, \dots, d_N)$ и $p(b|a, d_1, \dots, d_N)$

Задача 2

Решение 1

Необходимо посчитать соотношение:

$$p(a|d_1,\dots,d_N) = \frac{\sum_{b,\bar{c}} p(a,b,c_1,\dots,c_N,d_1,\dots,d_N)}{\sum_{a,b,\bar{c}} p(a,b,c_1,\dots,c_N,d_1,\dots,d_N)}$$

- Запишем $p(a|d_1,\dots,d_N) = p(d_1,\dots,d_N|a)p(a)/p(d_1,\dots,d_N)$.
- Вычислим $p(d_1,\dots,d_N|a)$.

$$p(d_1,\dots,d_N|a) = \sum_{c_1,\dots,c_N,b} p(b)\prod_{n=1}^{N} p(d_n|c_n)p(c_n|a,b) =$$

$$\sum_b p(b)\prod_{n=1}^{N}\left(\sum_{c_n} p(d_n|c_n)p(c_n|a,b)\right)$$

- $p(a)$ нам известен, $p(d_1,\dots,d_N) = \sum_a p(a)p(d_1,\dots,d_N)$

Задача 2

Решение 2

Необходимо посчитать соотношение:

$p(b|a, d_1, \dots, d_N) = \frac{\sum_{\bar{c}} p(a,b,c_1,\dots,c_N,d_1,\dots,d_N)}{\sum_{b,\bar{c}} p(a,b,c_1,\dots,c_N,d_1,\dots,d_N)}$

- Запишем $p(b|a, d_1, \dots, d_N) = p(a, d_1, \dots, d_N|b)p(b)/p(a, d_1, \dots, d_N)$.
- Вычислим $p(a, d_1, \dots, d_N|b)$.

$$p(a, d_1, \dots, d_N|b) = \sum_{c_1,\dots,c_N} p(a) \prod_{n=1}^{N} p(d_n|c_n)p(c_n|a, b) = p(a) \prod_{n=1}^{N} \left(\sum_{c_n} p(d_n|c_n)p(c_n|a, b) \right)$$

- $p(b)$ нам известен, $p(a, d_1, \dots, d_N) = \sum_b p(b)p(a, d_1, \dots, d_N|b)$

Задача 3

Модель бросания монеты

- Пусть $x_1, \dots, x_N \in \{0, 1\}$ — результаты подбрасывания монетки, выпадающей орлом с вероятностью q. Найти оценку максимального правдоподобия для q.
- Введем априорное распределение на параметр q:

$$p(q|\alpha, \beta) = \frac{q^{\alpha-1}(1-q)^{\beta-1}}{B(\alpha, \beta)}, \ \alpha, \beta > 0, \ Eq = \frac{\alpha}{\alpha+\beta};$$

$$p(x_1, \dots, x_N, q|\alpha, \beta) = p(q|\alpha, \beta) \prod_{i=1}^{N} p(x_i|q).$$

Найти распределение $p(q|x_1, \dots, x_N, \alpha, \beta)$.

Плотность бета-распределения

$$p(q|\alpha, \beta) \propto q^{\alpha-1}(1-q)^{\beta-1}, \ \alpha, \beta > 0$$

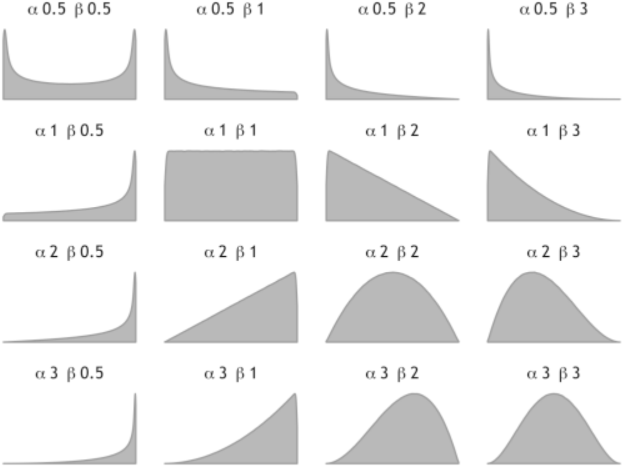


Иллюстрация найдена на matthewdharris.com

Задача 3

Решение

- Запишем логарифм функции правдоподобия: $L(q) = (\sum x_i) \log q + (N - \sum x_i) \log(1 - q)$. Приравнивая производную по q к нулю, получаем равенство

$$\frac{\mathrm{d}L}{\mathrm{d}q} = \sum x_i/q - (N - \sum x_i)/(1 - q) = 0$$

Решив уравнение, получаем ответ $\hat{q} = \frac{\sum x_i}{N}$

- Апостериорное распределение с точностью до множителя равно $q^{\alpha + \sum x_i - 1}(1 - q)^{\beta + N - \sum x_i - 1}$. Заметим, что это выражение соответствует ненормированной плотности бета-распределения с параметрами $\alpha' = \alpha + \sum x_i, \;\; \beta' = \beta + N - \sum x_i$

Задача 4

Точечные оценки параметров распределения

Зафиксируем функцию потерь L и будет выбирать параметр $\hat{\theta}$ согласно правилу

$$\hat{\theta} = \operatorname{argmin}_{\hat{\theta}} \mathbb{E}_{p(\theta|x_1,\ldots,x_N)} L(\theta, \hat{\theta}). \tag{1}$$

Какие оценки даст апостериорное распределение параметра монетки q из предыдущей задачи для

- квадратичной функции потерь $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$?
- модуля $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$?

Задача 4

Решение

Дифференциируем обе функции потерь по $\hat{\theta}$:

- $0 = \frac{\mathrm{d}}{\mathrm{d}\hat{\theta}}\mathbb{E}_p(\theta - \hat{\theta})^2 = 2\mathbb{E}_p\left(\theta - \hat{\theta}\right)$, следовательно $\hat{\theta} = \mathbb{E}_p\theta$
  Для выведенного в прошлом примере апостериорного распределения среднее будет равно $\frac{\alpha+\sum x_i}{\beta+N}$
- $0 = \frac{\mathrm{d}}{\mathrm{d}\hat{\theta}}\mathbb{E}_p|\theta - \hat{\theta}| = p(\theta < \hat{\theta}) - p(\theta \geq \hat{\theta})$, следовательно $\hat{\theta}$ совпадает с медианой распределения

Выбор априорного распределения

Обоснованностью называется функция правдоподобия для параметров модели:

$$p(x_1, \ldots, x_N|\alpha, \beta) = \int_{\mathbb{R}} \left[\prod_{i=1}^{N} p(x_i|q)\right] p(q|\alpha, \beta)\mathrm{d}q$$

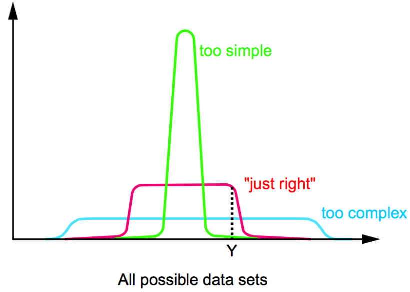


Рис.: Обоснованность трех моделей и разные данные

Метод наибольшей обоснованности предлагает выбирать априорное распределение в наибольшей обоснованностью

Задача 5

Подсчет обоснованности

Вычислите обоснованность $p(x_1, \ldots, x_N | \alpha, \beta)$ модели из предыдущей задачи

$$p(q|\alpha, \beta) = \frac{q^{\alpha-1}(1-q)^{\beta-1}}{B(\alpha, \beta)}, \;\; \alpha, \beta > 0$$

$$p(x_i|q) = q^{x_i}(1-q)^{1-x_i}$$

$$p(x_1, \ldots, x_N, q|\alpha, \beta) = p(q|\alpha, \beta) \prod_{i=1}^{N} p(x_i|q).$$

Задача 5

Решение

Воспользуемся определением бета-функции через интеграл $B(\alpha, \beta) = \int q^{\alpha-1}(1-q)^{\beta-1}\mathrm{d}q$

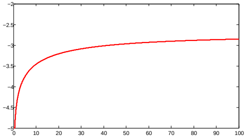
$$p(q|x_1, \dots, x_N) = \frac{1}{B(\alpha, \beta)} \int q^{\alpha + \sum x_i - 1}(1-q)^{\beta + N - \sum x_i - 1}\mathrm{d}q =$$

$$\frac{B(\alpha + \sum x_i, \alpha + N - \sum x_i)}{B(\alpha, \beta)}$$

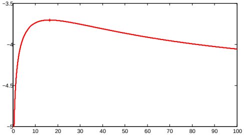
Оптимальное распределение в модели Beta-Binomial

Возьмем априорное распределение $q \sim Beta(\alpha, \alpha)$

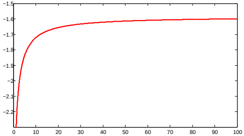
Построим значения обоснованности в зависимости от α



K=53, N=100



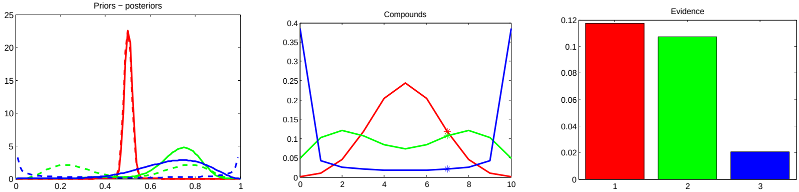
K=60, N=100



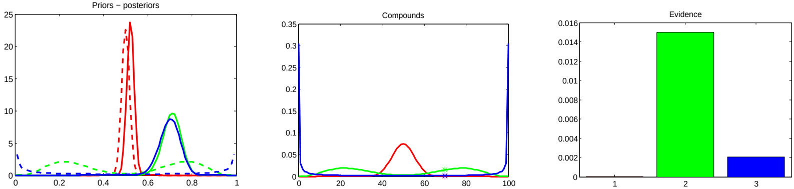
K=6, N=10

Обоснованность модели Betamix-Binomial

K=7, N=10:



K=70, N=100:



Столбцы:

1. априорное и апостериорное распределения
2. обоснованность в зависимости от K
3. обоснованность для заданного K

Задача 6

О репараметризации

- Пусть $\varepsilon \sim \mathcal{N}(0, I)$. Вычислить среднее и дисперсию случайного вектора $z = A\varepsilon + b$ для некоторой матрицы A и вектора b.
- Пусть случайный вектор z имеет нормальное распределение

$$\mathcal{N}(z|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(z-\mu)^T \Sigma^{-1} (z-\mu)\right)$$

  Для каких пар (A, b) вектор z представим в виде $z = A\varepsilon + b$?
- Какова сложность поиска такого представления?

Задача 6

Решение

- $z \sim \mathcal{N}(b, A \cdot A^T)$
  - $\mathbb{E}_z z = \mathbb{E}_\varepsilon A\varepsilon + b = 0 + b$
  - $\mathbb{E}_z (z-b)(z-b)^T = \mathbb{E}_\varepsilon (A\varepsilon)(A\varepsilon)^T = A\mathbb{E}_\varepsilon \varepsilon\varepsilon^T A^T = A \cdot A^T$
- Найдем разложение Холецкого матрицы ковариации: $\Sigma = L \cdot L^T$. Случайный вектор $L\varepsilon + \mu$ имеет распределение $\mathcal{N}(\mu, \Sigma)$

Задача 7

Байесовская линейная регрессия

Дано

- Размеченная выборка из N объектов $(X, y) = \{x_i, y_i\}_{i=1}^N$
- Байесовская модель линейной регрессии:

$$
\begin{aligned}
p(y, w | X, \alpha, \beta) &= p(y | w, X, \beta) p(w | \alpha), \\
p(y | w, X, \beta) &= \mathcal{N}(y | Xw, \beta^{-1} I), \\
p(w | \alpha) &= \mathcal{N}(w | 0, \alpha^{-1} I).
\end{aligned}
$$

Найти

Апостериорное распределение на веса $p(w | X, y, \alpha, \beta)$

Задача 7

Решение, слайд 1

- Заметим, что апостериорное распределение лежит в классе нормальных

$$p(w|X,y) \propto p(w)p(y|X,w) \propto$$
$$\exp\left[-\frac{\alpha}{2}w^T w - \frac{\beta}{2}(y - Xw)^T(y - Xw),\right]$$

  поскольку его плотность - экспонента от квадратичной функции от w

- Будем искать параметры (A, b) апостериорного распределения $p(y|X,\alpha,\beta) = \mathcal{N}(b, A^{-1})$

Задача 7

Решение, слайд 2

- Для формы $-\frac{1}{2}(w-b)^T A(w-b)$ с симметрической матрицей A выполнено

$$\frac{\partial}{\partial w}\left(-\frac{1}{2}(w-b)^T A(w-b)\right) = -A(w-b)$$
$$\frac{\partial^2}{\partial w^2}\left(-\frac{1}{2}(w-b)^T A(w-b)\right) = -A$$

- Дифференцируя $-\frac{\alpha}{2}w^T w - \frac{\beta}{2}(y-Xw)^T(y-Xw)$, получим

$$-A(w-b) = -\alpha w + \beta(y-Xw)^T X$$
$$-A = -\alpha I - \beta X^T \cdot X$$

Задача 7

Решение, слайд 3

- Дифференциируя $-\frac{\alpha}{2}w^T w - \frac{\beta}{2}(y - Xw)^T(y - Xw)$, получим

$$
\begin{aligned}
-A(w-b) &= -\alpha w + \beta(y - Xw)^T X \\
-A &= -\alpha I - \beta X^T \cdot X
\end{aligned}
$$

- Из второго соотношения получаем выражение для обратной матрицы ковариаций: $A = \alpha I + \beta X^T \cdot X$
- Соотношение между свободным членами в первом равенстве дает $Ab = \beta yX$, отсюда $b = \beta A^{-1} yX$

Задача 8

Подсчет обоснованности

Дано

- Размеченная выборка из N объектов $(X, y) = \{x_i, y_i\}_{i=1}^N$
- Байесовская модель линейной регрессии:

$$
\begin{aligned}
p(y, w|X, \alpha, \beta) &= p(y|w, X, \beta)p(w|\alpha),\\
p(y|w, X, \beta) &= \mathcal{N}(y|Xw, \beta I),\\
p(w|\alpha) &= \mathcal{N}(w|0, \alpha I).
\end{aligned}
$$

Найти

Выражение для обоснованности модели $p(y|X, \alpha, \beta)$

Задача 8

Решение

- Вектор y представим в виде суму двух независимых нормальных случайных величин:

$$
\begin{aligned}
y &= Xw + b + \varepsilon \\
w &\sim \mathcal{N}(0, \alpha^{-1} I) \\
\varepsilon &\sim \mathcal{N}(0, \beta^{-1} I)
\end{aligned}
$$

- Вычислим среднее и матрицу ковариаций y:

$$
\begin{aligned}
&\mathbb{E}_y y = \mathbb{E}_{w,\varepsilon} [Xw + b + \varepsilon] = X \cdot 0 + b + 0 \\
&\mathbb{E}_y (y - b)(y - b)^T = \mathbb{E}_{w,\varepsilon} (Xw + \varepsilon)(Xw + \varepsilon)^T \\
&= X \mathbb{E}_w w w^T X^T + 2\mathbb{E}_{w,\varepsilon} \varepsilon w^T X + \mathbb{E}_\varepsilon \varepsilon \varepsilon^T \\
&= \alpha^{-1} X X^T + \beta^{-1} I
\end{aligned}
$$

- Итого $p(y|X, \alpha, \beta) = \mathcal{N}(b, \beta^{-1} I + \alpha^{-1} X X^T)$

Модель метода релевантных векторов

Метод релевантных векторов вводит более богатое семейство априорных распределений на веса модели, это соответствует индивидуальным коэффициентам регуляризации для всех весов модели.

$$p(y, w|X, A, \beta) = p(y|w, X, \beta)p(w|A),$$
$$p(y|w, X, \beta) = \mathcal{N}(y|Xw, \beta^{-1}I),$$
$$p(w|\alpha) = \mathcal{N}(w|0, A), \; A = \mathrm{diag}(\alpha_1^{-1}, \dots, \alpha_d^{-1})$$

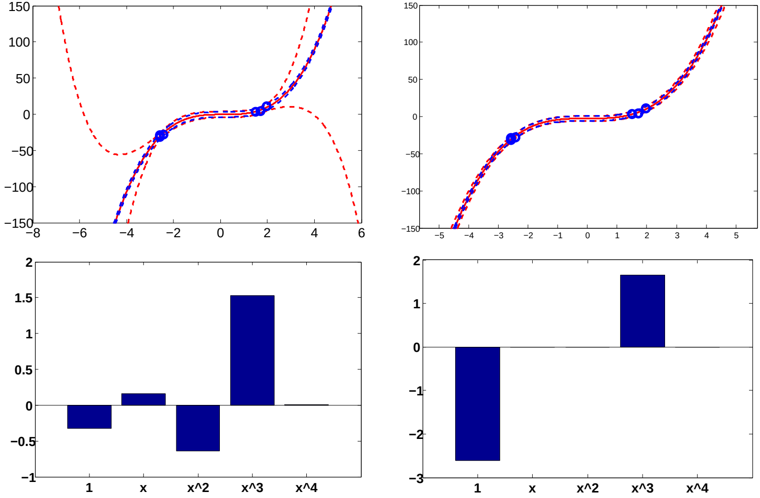
Параметры α находятся на основании принципа наибольше обоснованности:

$$p(y|X, \bar{\alpha}, \beta) \to \max_{\bar{\alpha}}$$

Метод релевантных векторов

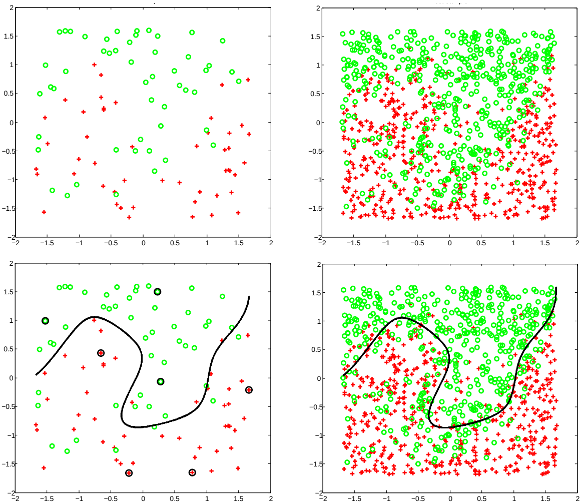
Полиномиальная регрессия

$$w \sim \mathcal{N}(0, \alpha I) \qquad\qquad w \sim \mathcal{N}(0, \mathrm{diag}(\alpha_1, \ldots, \alpha_d))$$



Метод релевантных векторов

Классификация, RBF признаки



Метод релевантных векторов

Классификация на 4 класса, RBF признаки