

Задание 3 по курсу "Байесовский выбор модели"

Грабовой Андрей, группа 574

Задача 1

По условию задачи:

$$p(\mathbf{X}, \mathbf{y}, \mathbf{w} | \mathbf{A}) = \prod_i N(\mathbf{x}_i | \mathbf{0}, \sigma^2 \mathbf{I}_n) N(\mathbf{w} | \mathbf{0}, \mathbf{A}^{-1}) \prod_j p(y_j | \mathbf{x}_j, \mathbf{w}), \quad (1.1)$$

где $p(y_j = 1 | \mathbf{x}_j, \mathbf{w}) = \frac{1}{1 + \exp(-\mathbf{w}^\top \mathbf{x}_j)}$

Для простоты запишем (1.1) в следующем общем виде:

$$p(\mathbf{X}, \mathbf{y}, \mathbf{w} | \mathbf{A}) = p(\mathbf{X}) p(\mathbf{w} | \mathbf{A}) p(\mathbf{y} | \mathbf{X}, \mathbf{w}). \quad (1.2)$$

а) По формуле Байеса:

$$p(\mathbf{w} | \mathbf{X}, \mathbf{y}, \mathbf{A}) = \frac{p(\mathbf{y} | \mathbf{X}, \mathbf{w}) p(\mathbf{w} | \mathbf{A})}{\int_{\mathbf{w}' \in \mathbb{R}^n} p(\mathbf{y} | \mathbf{X}, \mathbf{w}') p(\mathbf{w}' | \mathbf{A}) d\mathbf{w}'} = \frac{\mathcal{Q}(\mathbf{w})}{\int_{\mathbf{w}' \in \mathbb{R}^n} \mathcal{Q}(\mathbf{w}')}, \quad (1.3)$$

где введено обозначение $\mathcal{Q}(\mathbf{w}) = p(\mathbf{y} | \mathbf{X}, \mathbf{w}) p(\mathbf{w} | \mathbf{A})$.

Выполним аппроксимацию Лапласа:

$$\begin{aligned} \log \mathcal{Q}(\mathbf{w}) &\approx \log \mathcal{Q}(\mathbf{w}_{\text{MAP}}) + \cancel{\nabla \log \mathcal{Q}(\mathbf{w}_{\text{MAP}})} + \frac{1}{2} (\mathbf{w} - \mathbf{w}_{\text{MAP}})^\top \nabla \nabla^\top \log \mathcal{Q}(\mathbf{w}_{\text{MAP}}) (\mathbf{w} - \mathbf{w}_{\text{MAP}}) = \\ &= \log \mathcal{Q}(\mathbf{w}_{\text{MAP}}) - \frac{1}{2} (\mathbf{w} - \mathbf{w}_{\text{MAP}})^\top \mathbf{H}^{-1} (\mathbf{w} - \mathbf{w}_{\text{MAP}}), \end{aligned} \quad (1.4)$$

где введено обозначение $\mathbf{H}^{-1} = -\nabla \nabla^\top \log \mathcal{Q}(\mathbf{w}_{\text{MAP}})$.

Для нашей задачи найдем \mathbf{H}^{-1} :

$$\begin{aligned} \mathbf{H}^{-1} &= -\nabla \nabla^\top \left(-\frac{1}{2} \mathbf{w}^\top \mathbf{A}^{-1} \mathbf{w} - \mathbf{1}^\top \log(1 + \exp(-\mathbf{X}^\top \mathbf{w})) \right) = \\ &= \mathbf{A}^{-1} + \nabla \nabla^\top \mathbf{1}^\top \log(1 + \exp(-\mathbf{X}^\top \mathbf{w})) = \\ &= \mathbf{A}^{-1} + \sum_{i=1}^m \nabla \nabla^\top \log(1 + \exp(-\mathbf{x}_i^\top \mathbf{w})) = \end{aligned}$$

$$= \mathbf{A}^{-1} + \sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i^T \frac{\exp(-\mathbf{x}_i^T \mathbf{w})}{1 + \exp(-\mathbf{x}_i^T \mathbf{w})} - \mathbf{x}_i \mathbf{x}_i^T \frac{\exp(-2\mathbf{x}_i^T \mathbf{w})}{(1 + \exp(-\mathbf{x}_i^T \mathbf{w}))^2}. \quad (1.5)$$

Тогда получаем:

$$\mathcal{Q}(\mathbf{w}) \approx \mathcal{Q}(\mathbf{w}_{\text{MAP}}) \exp \left(-\frac{1}{2} (\mathbf{w} - \mathbf{w}_{\text{MAP}})^T \mathbf{H}^{-1} (\mathbf{w} - \mathbf{w}_{\text{MAP}}) \right). \quad (1.6)$$

Подставляя (1.6) в (1.3) получим:

$$\begin{aligned} p(\mathbf{w}|\mathbf{X}, \mathbf{y}, \mathbf{A}) &\approx \frac{\mathcal{Q}(\mathbf{w}_{\text{MAP}}) \exp \left(-\frac{1}{2} (\mathbf{w} - \mathbf{w}_{\text{MAP}})^T \mathbf{H}^{-1} (\mathbf{w} - \mathbf{w}_{\text{MAP}}) \right)}{\int_{\mathbf{w}' \in \mathbb{R}^n} \mathcal{Q}(\mathbf{w}_{\text{MAP}}) \exp \left(-\frac{1}{2} (\mathbf{w}' - \mathbf{w}_{\text{MAP}})^T \mathbf{H}^{-1} (\mathbf{w}' - \mathbf{w}_{\text{MAP}}) \right) d\mathbf{w}'} = \\ &= \frac{\exp \left(-\frac{1}{2} (\mathbf{w} - \mathbf{w}_{\text{MAP}})^T \mathbf{H}^{-1} (\mathbf{w} - \mathbf{w}_{\text{MAP}}) \right)}{\int_{\mathbf{w}' \in \mathbb{R}^n} \exp \left(-\frac{1}{2} (\mathbf{w}' - \mathbf{w}_{\text{MAP}})^T \mathbf{H}^{-1} (\mathbf{w}' - \mathbf{w}_{\text{MAP}}) \right) d\mathbf{w}'} = \\ &= N(\mathbf{w}_{\text{MAP}}, \mathbf{H}). \end{aligned} \quad (1.7)$$

Оценим \mathbf{w}_{MAP} :

$$\mathbf{w}_{\text{MAP}} = \arg \max_{\mathbf{w} \in \mathbb{R}^n} p(\mathbf{w}|\mathbf{X}, \mathbf{y}, \mathbf{A}) = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \{-\log p(\mathbf{y}|\mathbf{X}, \mathbf{w}) - \log p(\mathbf{w}|\mathbf{A})\}, \quad (1.8)$$

где

$$p(\mathbf{y}|\mathbf{X}, \mathbf{w}) = \prod_i \hat{p}_i^{y_i} (1 - \hat{p}_i)^{1-y_i}; \quad -\log p(\mathbf{w}|\mathbf{A}) = \frac{1}{2} \mathbf{w}^T \mathbf{A}^{-1} \mathbf{w}; \quad \hat{\mathbf{p}} = \frac{1}{1 + \exp(-\mathbf{X}^T \mathbf{w})}, \quad (1.9)$$

Подставляя (1.9) в (1.8) получаем:

$$\begin{aligned} \mathbf{w}_{\text{MAP}} &= \arg \min_{\mathbf{w} \in \mathbb{R}^n} \{-\mathbf{y}^T \log \hat{\mathbf{p}} - (1 - \mathbf{y})^T \log(1 - \hat{\mathbf{p}}) + \frac{1}{2} \mathbf{w}^T \mathbf{A}^{-1} \mathbf{w}\} = \\ &= \arg \min_{\mathbf{w} \in \mathbb{R}^n} \left\{ -\mathbf{y}^T \log \frac{1}{1 + \exp(-\mathbf{X}^T \mathbf{w})} - (1 - \mathbf{y})^T \log \frac{\exp(-\mathbf{X}^T \mathbf{w})}{1 + \exp(-\mathbf{X}^T \mathbf{w})} + \frac{1}{2} \mathbf{w}^T \mathbf{A}^{-1} \mathbf{w} \right\} = \\ &= \arg \min_{\mathbf{w} \in \mathbb{R}^n} \left\{ (1 - \mathbf{y})^T \mathbf{X}^T \mathbf{w} + \mathbf{1}^T \log(1 + \exp(-\mathbf{X}^T \mathbf{w})) + \frac{1}{2} \mathbf{w}^T \mathbf{A}^{-1} \mathbf{w} \right\}, \end{aligned} \quad (1.10)$$

где введя обозначения $\mathcal{L}(\mathbf{w}|\mathbf{X}, \mathbf{y}, \mathbf{A}) = (1 - \mathbf{y})^T \mathbf{X}^T \mathbf{w} + \mathbf{1}^T \log(1 + \exp(-\mathbf{X}^T \mathbf{w})) + \frac{1}{2} \mathbf{w}^T \mathbf{A}^{-1} \mathbf{w}$ получим следующую оптимизационную задачу для нахождения \mathbf{w}_{MAP} :

$$\mathbf{w}_{\text{MAP}} = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \mathcal{L}(\mathbf{w}|\mathbf{X}, \mathbf{y}, \mathbf{A}). \quad (1.11)$$

Новая версия

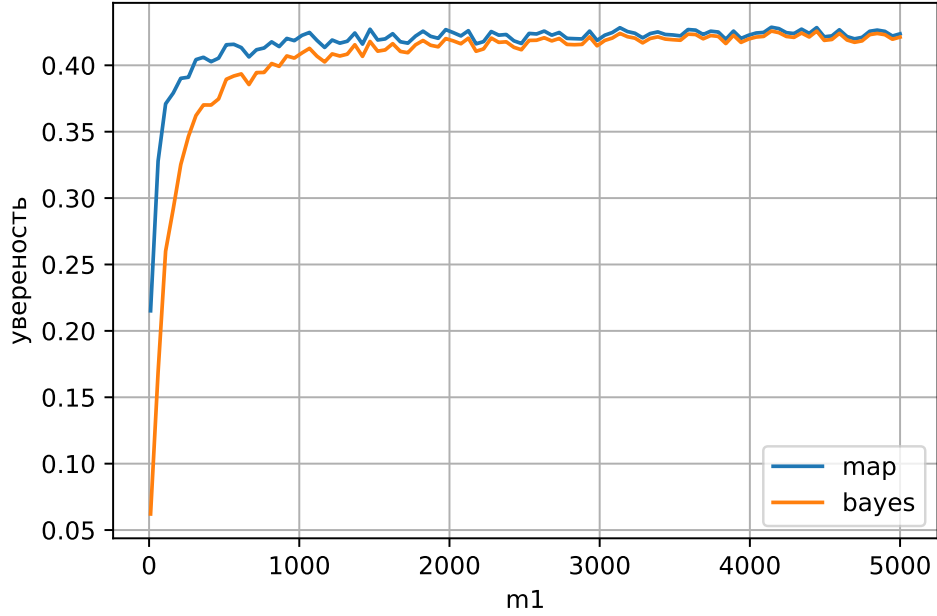


Рис. 1: График уверенности, от размера обучающей выборки

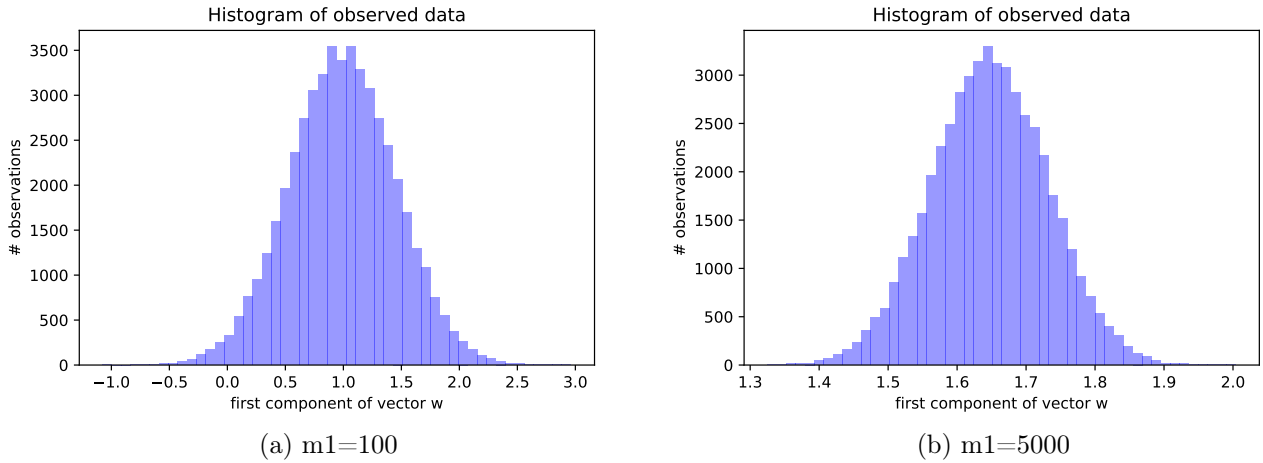


Рис. 2: Гистограмма распределения первой компоненты вектора $\mathbf{w} \sim N(\mathbf{w}_{\text{MAP}}, \mathbf{H})$

b) Пусть имеем следующие обучающие и тестовые выборки $(\mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}}) \in (\mathbb{R}^{m_1 \times n}, \{0, 1\}^{m_1})$ и $(\mathbf{X}_{\text{test}}, \mathbf{y}_{\text{test}}) \in (\mathbb{R}^{m_2 \times n}, \{0, 1\}^{m_2})$. Пусть $m_2 = 1000$, $\sigma^2 = 1$, $\mathbf{A} = \mathbf{I}_n$, $n = 50$.

Пусть $\hat{\mathbf{p}}$ — вектор оценок вероятностей принадлежности объекта к классу 1 для некоторого классификатора на некоторой выборке \mathbf{X} .

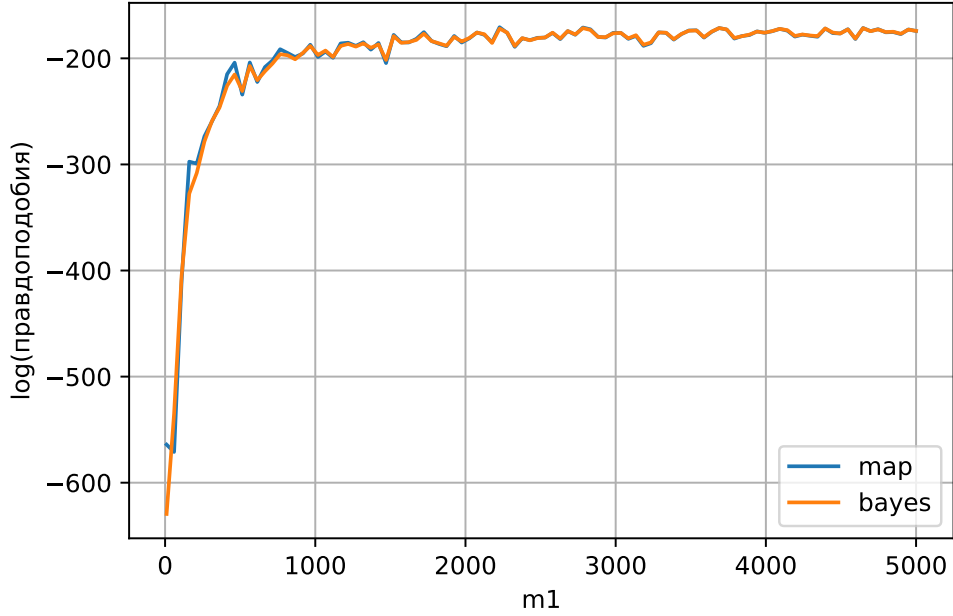


Рис. 3: График логарифма правдоподобия, от размера обучающей выборки

Введем уверенность $C(\hat{\mathbf{p}})$ классификатора:

$$C(\hat{\mathbf{p}}) = \sum_{i=1}^{m_2} |\hat{p}_i - 0.5|, \quad (1.12)$$

где m_2 — размер выборки.

Рассмотрим правдоподобие выборки для заданного вектора $\hat{\mathbf{p}}$ и истинных \mathbf{y} :

$$l(\mathbf{y}, \hat{\mathbf{p}}) = \prod_{i=1}^{m_2} \hat{p}_i^{y_i} (1 - \hat{p}_i)^{1-y_i}. \quad (1.13)$$

Сравним уверенность и правдоподобие для точечного МАР-классификатора:

$$\hat{\mathbf{p}}_{test}^{MAP} = \frac{1}{1 + \exp(-\mathbf{X}^T \mathbf{w}_{MAP})}, \quad (1.14)$$

и для байесовского классификатора:

$$\hat{\mathbf{p}}_{test}^{BAYES} = \int_{\mathbf{w} \in \mathbb{R}^n} \frac{p(\mathbf{w} | \mathbf{X}_{train}, \mathbf{y}_{train}, \mathbf{A})}{1 + \exp(-\mathbf{X}^T \mathbf{w})} d\mathbf{w} = \mathbb{E}_{N(\mathbf{w}_{MAP}, \mathbf{H})} \frac{1}{1 + \exp(-\mathbf{X}^T \mathbf{w})}. \quad (1.15)$$

В виду что в формуле (1.15) сложный интеграл его будем считать сэмплированием (метод Монте Карло).

Выборка генерировалась следующим образом. Сначала генерировался некоторый $\mathbf{w} \sim N(\mathbf{0}, \mathbf{A})$. После чего генерировалась выборка $\mathbf{X}_{test} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ — единая для всего эксперимента. После этого генерировался вектор ответов $\mathbf{y}_{test} = \text{Be}\left(\frac{1}{1+\exp(-\mathbf{X}^\top \mathbf{w})}\right)$. Далее для каждого m_1 генерировались выборки $\mathbf{X}_{train}, \mathbf{y}_{train}$.

На рис. 1 показано, что точечный МАР-классификатор является более уверенным в отличии от байесовского классификатора, что следует из того, что байесовский классификатор усредняет свой ответ. Важно заметить, что при увеличении размера выборки они ведут себя почти одинаково, это следует из того, что при большом количестве данных у нас становится все более выраженный пик распределения. Изменения разброса значений \mathbf{w} показано на рис. 2. Видно, что при $m_1 = 100$ разброс в первой компоненте намного больше чем разброс в первой компоненте при $m_1 = 5000$.

На рис. 3 показано, что байесовский классификатор и МАР-классификатор дают более менее одинаковое правдоподобие модели.

Из описанного выше можно сделать вывод, что байесовский классификатор является более устойчивым и при большом количестве данных он сходиться к распределению который имеет пик в матожидание, если этот вектор хорошо описывает выборку.

Задача 2

Пусть имеется модель линейной регрессии:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (2.1)$$

где σ^2 — задано, и задано априорное распределение $p(\mathbf{w}|\mathbf{m}, \text{diag}(\mathbf{s}))$.

Из (2.1) получаем распределение на $p(\mathbf{y}, \mathbf{w}|\mathbf{X}, \mathbf{m}, \mathbf{s})$:

$$p(\mathbf{y}, \mathbf{w}|\mathbf{X}, \mathbf{m}, \mathbf{s}) = p(\mathbf{y}|\mathbf{X}, \mathbf{w})p(\mathbf{w}|\mathbf{m}, \mathbf{s}) = N(\mathbf{X}\mathbf{w}, \sigma^2 \mathbf{I})N(\mathbf{m}, \text{diag}(\mathbf{s})). \quad (2.2)$$

Тогда по формуле Байеса получаем апостериорное распределение на вектор \mathbf{w} :

$$p(\mathbf{w}|\mathbf{X}, \mathbf{y}, \mathbf{m}, \mathbf{s}) = \frac{p(\mathbf{y}, \mathbf{w}|\mathbf{X}, \mathbf{m}, \mathbf{s})}{p(\mathbf{y}|\mathbf{X}, \mathbf{m}, \mathbf{s})} = \frac{p(\mathbf{y}|\mathbf{X}, \mathbf{w})p(\mathbf{w}|\mathbf{m}, \mathbf{s})}{\int_{\mathbf{w}' \in \mathbb{R}^n} p(\mathbf{y}, \mathbf{w}'|\mathbf{X}, \mathbf{m}, \mathbf{s})d\mathbf{w}'} = \frac{\mathcal{Q}(\mathbf{w})}{\int_{\mathbf{w}' \in \mathbb{R}^n} \mathcal{Q}(\mathbf{w}')d\mathbf{w}'}, \quad (2.3)$$

где введено обозначение $\mathcal{Q}(\mathbf{w}) = p(\mathbf{y}|\mathbf{X}, \mathbf{w})p(\mathbf{w}|\mathbf{m}, \mathbf{s})$.

Найдем $\mathbf{w}_{\text{МАР}}$:

$$\begin{aligned} \mathbf{w}_{\text{МАР}} &= \arg \max_{\mathbf{w} \in \mathbb{R}^n} \mathcal{Q}(\mathbf{w}) = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\mathbf{w})^\top (\mathbf{y} - \mathbf{X}\mathbf{w}) + \frac{1}{2}(\mathbf{w} - \mathbf{m})^\top \mathbf{A}^{-1}(\mathbf{w} - \mathbf{m}) = \\ &= \arg \min_{\mathbf{w} \in \mathbb{R}^n} \mathcal{L}(\mathbf{w}), \end{aligned} \quad (2.4)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{w}^\top} &= -\frac{1}{\sigma^2} \mathbf{X}^\top (\mathbf{y} - \mathbf{X}\mathbf{w}) + \mathbf{A}^{-1}(\mathbf{w} - \mathbf{m}) = 0 \Rightarrow \\ &\Rightarrow -\frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{y} + \left(\frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X} + \mathbf{A}^{-1} \right) \mathbf{w} - \mathbf{A}^{-1} \mathbf{m} = 0 \Rightarrow \end{aligned}$$

$$\Rightarrow \mathbf{w}_{\text{MAP}} = \left(\frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X} + \mathbf{A}^{-1} \right)^{-1} \left(\frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{y} + \mathbf{A}^{-1} \mathbf{m} \right), \quad (2.5)$$

где $\mathbf{A} = \text{diag}(\mathbf{s})$

Найдем \mathbf{H} :

$$\mathbf{H} = (\nabla \nabla^\top \mathcal{L})^{-1} = \left(\frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X} + \mathbf{A}^{-1} \right)^{-1}, \quad (2.6)$$

Получаем:

$$p(\mathbf{w} | \mathbf{X}, \mathbf{y}, \mathbf{m}, \mathbf{s}) = N(\mathbf{w}_{\text{MAP}}, \mathbf{H}). \quad (2.7)$$

Если некоторое $s_i = 0$, то как видно из формул (2.5) и (2.6) получим, что апостериорное распределение i -го элемента вектора равно константе m_i .

Если же воспользоваться свойством сопряженности, то получится такие же параметры нормального распределения, я честно пересчитал, но сюда не стал вносить. Это достаточно очевидно, потому что на самом деле мы просто разложили квадратичную функцию до второго порядка, это мы и получили саму функцию. То есть в данном случае аппроксимация Лапласа дает точный ответ.

Новая версия

$$\hat{\mathbf{m}}, \hat{\mathbf{s}} = \arg \max_{\mathbf{m}, \mathbf{s} \in \mathbb{R}^n} \mathbb{E}_{N(\mathbf{w} | \mathbf{w}_{\text{MAP}}, \mathbf{H})} \log p(\mathbf{w}, \mathbf{y} | \mathbf{X}, \mathbf{m}, \mathbf{s}, \sigma) = \arg \max_{\mathbf{m}, \mathbf{s} \in \mathbb{R}^n} \mathcal{F}(\mathbf{m}, \mathbf{s}). \quad (2.13)$$

Используя формулу (2.2) получаем:

$$\begin{aligned} \mathcal{F}(\mathbf{m}, \mathbf{s}) &= \mathbb{E}_{N(\mathbf{w} | \mathbf{w}_{\text{MAP}}, \mathbf{H})} \log (N(\mathbf{y} | \mathbf{X} \mathbf{w}, \sigma^2 \mathbf{I}) N(\mathbf{w} | \mathbf{m}, \mathbf{s})) = \\ &= \mathbb{E}_{N(\mathbf{w} | \mathbf{w}_{\text{MAP}}, \mathbf{H})} \log N(\mathbf{y} | \mathbf{X} \mathbf{w}, \sigma^2 \mathbf{I}) + \mathbb{E}_{N(\mathbf{w} | \mathbf{w}_{\text{MAP}}, \mathbf{H})} \log N(\mathbf{w} | \mathbf{m}, \mathbf{s}) = \\ &= C - \frac{1}{2} \sum_{i=1}^n \log s_i - \frac{1}{2\sigma^2} \mathbb{E}_{N(\mathbf{w} | \mathbf{w}_{\text{MAP}}, \mathbf{H})} \|\mathbf{y} - \mathbf{X} \mathbf{w}\|^2 - \frac{1}{2} \mathbb{E}_{N(\mathbf{w} | \mathbf{w}_{\text{MAP}}, \mathbf{H})} (\mathbf{w} - \mathbf{m})^\top \mathbf{A}^{-1} (\mathbf{w} - \mathbf{m}) = \\ &= C - \frac{1}{2} \sum_{i=1}^n \log s_i - \frac{1}{2\sigma^2} \mathbb{E}_{N(\mathbf{w} | \mathbf{w}_{\text{MAP}}, \mathbf{H})} \|\mathbf{y} - \mathbf{X} \mathbf{w}\|^2 - \frac{1}{2} \sum_{i=1}^n \frac{\mathbb{E}(w_i - m_i)^2}{s_i}. \end{aligned} \quad (2.14)$$

Если мы считаем, что на этом шаге у нас \mathbf{w}_{MAP} постоянная, тогда производная заносится под знак матожидания и мы получаем следующий результат:

$$\frac{\partial \mathcal{F}}{\partial \mathbf{m}^\top} = \mathbf{A}^{-1} \mathbf{m} - \mathbb{E}_{N(\mathbf{w} | \mathbf{w}_{\text{MAP}}, \mathbf{H})} \mathbf{A}^{-1} \mathbf{w} = \mathbf{0} \Rightarrow \mathbf{m} = \mathbf{w}_{\text{MAP}}, \quad (2.15)$$

$$\frac{\partial \mathcal{F}}{\partial s_k} = -\frac{1}{2s_k} + \frac{1}{2s_k^2} \mathbb{E}_{N(\mathbf{w} | \mathbf{m}, \mathbf{H})} (w_k - m_k)^2 = 0 \Rightarrow s_k = D w_k. \quad (2.16)$$

Из формул (2.5) и (2.15) получаем:

$$\mathbf{m} = \left(\frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X} + \mathbf{A}^{-1} \right)^{-1} \left(\frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{y} + \mathbf{A}^{-1} \mathbf{m} \right) \Rightarrow \mathbf{m} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}, \quad (2.17)$$

то есть максимум обоснованности достигается, когда \mathbf{w} имеет распределение с матожиданием которое является решением уравнения $\mathbf{y} = \mathbf{X}\mathbf{m}$ методом наименьших квадратов.

Из формул (2.6) и (2.16) получаем:

$$\text{diag}(\mathbf{A}) = \text{diag} \left[\left(\frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X} + \mathbf{A}^{-1} \right)^{-1} \right], \quad (2.18)$$

явно проитерировать это не смог, поэтому оставил в такой рекуррентной формуле.

Задача 3

Пусть имеются две двухсторонние монеты, случайно и независимо выбранные из всех существующих монет достоинством в 2 рубля. Пусть было произведено $n_1 = 10$ бросаний первой монеты и $n_2 = 10000$ бросаний второй. Среди n_1 результатов бросания первой монеты было $k_1 = 3$ орла, а среди n_2 бросаний второй — $k_2 = 5050$ орлов.

Новая версия

а) Пусть $k \in \mathbb{N}_0$ — количество выпавших орлов, тогда имеем следующую вероятностную модель эксперимента:

$$p(k_1, p_1, k_2, p_2 | n_1, n_2) = q(p_1)q(p_2)\text{Bin}(k_1 | p_1, n_1)\text{Bin}(k_2 | p_2, n_2);$$

$$q(p) = \begin{cases} 1 & p \in [0, 1] \\ 0, & p \notin [0, 1] \end{cases}, \quad (3.17)$$

где $q(p)$ — априорное распределение вероятности выпадении орла на монете. Априорное распределение вероятности задано таким образом, что все монеты в мире равновероятны (теперь сделал так, чтобы потом интеграл хорошо брался, как-раз сразу видно сопряженное распределение).

б) По формуле Байеса получим апостериорное распределение $q(p | k, n)$:

$$q(p | k, n) = \frac{p(k, p | n)}{p(k | n)} = \frac{C_n^k p^k (1-p)^{n-k}}{\int_0^1 C_n^k p'^k (1-p')^{n-k} dp'} = B(p | k+1, n-k+1), \quad (3.18)$$

где $B(\alpha, \beta)$ — бета распределение.

с) Рассмотрим следующие модели M_1 и M_2 :

$$p_{M_1}(k_1, p_1, k_2, p_2 | n_1, n_2) = p_{M_1}(k_1, p, k_2 | n_1, n_2) = q(p)\text{Bin}(k_1 | p, n_1)\text{Bin}(k_2 | p, n_2), \quad (3.19)$$

$$p_{M_2}(k_1, p_1, k_2, p_2 | n_1, n_2) = q(p_1)q(p_2)\text{Bin}(k_1 | p_1, n_1)\text{Bin}(k_2 | p_2, n_2), \quad (3.20)$$

где $p(M_1) = p(M_2) = \frac{1}{2}$.

Тогда общая модель выглядит следующим образом:

$$p(k_1, p_1, k_2, p_2, M_i | n_1, n_2) = p(M_i) p_{M_i}(k_1, p_1, k_2, p_2 | n_1, n_2). \quad (3.21)$$

Найдем апостериорные вероятности моделей:

$$\begin{aligned} p(M_i | k_1, n_1, k_2, n_2) &= \frac{p(M_i, k_1, k_2 | n_1, n_2)}{p(M_1, k_1, k_2 | n_1, n_2) + p(M_2, k_1, k_2 | n_1, n_2)} = \\ &= \frac{p(M_i) p_{M_i}(k_1, k_2 | n_1, n_2)}{p(M_1) p_{M_1}(k_1, k_2 | n_1, n_2) + p(M_2) p_{M_2}(k_1, k_2 | n_1, n_2)}. \end{aligned} \quad (3.22)$$

Найдем $p_{M_1}(k_1, k_2 | n_1, n_2)$ и $p_{M_2}(k_1, k_2 | n_1, n_2)$:

$$p_{M_1}(k_1, k_2 | n_1, n_2) = C \int_0^1 p^{k_1+k_2} (1-p)^{n_2+n_1-k_2-k_1} dp, \quad (3.23)$$

$$p_{M_2}(k_1, k_2 | n_1, n_2) = C \int_0^1 \int_0^1 p_1^{k_1} (1-p_1)^{n_1-k_1} p_2^{k_2} (1-p_2)^{n_2-k_2} dp_1 dp_2. \quad (3.24)$$

Получаем следующие апостериорные вероятности моделей:

$$p(M_1 | k_1, n_1, k_2, n_2) = \frac{Beta(\alpha_3, \beta_3)}{Beta(\alpha_3, \beta_3) + Beta(\alpha_1, \beta_1) Beta(\alpha_2, \beta_2)}, \quad (3.25)$$

где $(\alpha_{1,2}, \beta_{1,2}) = (k_{1,2} + 1, n_{1,2} - k_{1,2} + 1)$; $(\alpha_3, \beta_3) = (k_1 + k_2 + 1, n_1 + n_2 - k_1 - k_2 + 1)$.

Посчитав, получаем следующие вероятности:

$$p(M_1 | k_1, n_1, k_2, n_2) = 0.55, \quad p(M_2 | k_1, n_1, k_2, n_2) = 0.45. \quad (3.26)$$

Снова таки, апостериорная вероятность более простой модели больше чем более сложной.