

Отчёт по практическому заданию №2. Байесовская смесь распределений Бернулли

Ю. Н. Лукашкина

julialukashkina@gmail.ru

МГУ имени М. В. Ломоносова, Москва

28 ноября 2014 г.

Аннотация

Данный документ содержит отчет по заданию БММО группы 417 "Байесовская смесь распределений Бернулли".

Содержание

1 Введение	1
2 Вывод формул	1
2.1 Описание модели	1
2.2 Е-шаг. Формулы для пересчета $q(\mu), q(Z)$	2
2.3 М-шаг. Формула для пересчёта π	3
2.4 Формула для подсчета вариационной нижней границы	4
3 Модельные данные	4
4 Данные MNIST	4
4.1 Функционал $\mathcal{L}(q)$	5
4.2 Зависимость числа получаемых кластеров от априорного значения колличества компонент	5
4.3 Зависимость числа получаемых кластеров от α	6
4.4 Зависимость числа получаемых кластеров от a, b	6
4.5 Визуализация полученных центров кластеров $\mathbb{E}_{q(\mu)}\mu$	6
5 Исследование логарифма правдоподобия	6
6 Классификация с новым признаком для $n_{\text{ого}}$ объекта: величины $q(z_{nk} = 1)$. Матрица точно-сти.	7
7 Выводы	8

1 Введение

В данном задании было предложено вывести формулы для вариационного ЕМ-алгоритма, реализовать данный ЕМ-алгоритм, посмотреть как изменяется кластеризация в зависимости от подаваемых параметров, произвести классификацию с помощью новых признаков, получаемых после работы ЕМ-алгоритма.

2 Вывод формул

2.1 Описание модели

Пусть $\mathbf{x} = (x_1, \dots, x_D)^T$ - набор из D бинарных случайных величин x_i , каждая из которых имеет распределение Бернулли с параметром μ_i . Таким образом

$$p(\mathbf{x} | \mu) = \prod_{i=1}^D \mu_i^{x_i} (1 - \mu_i)^{(1-x_i)},$$

где $\mu = (\mu_1, \dots, \mu_D)^T$. Рассмотрим смесь таких распределений:

$$p(\mathbf{x} | \mu, \pi) = \sum_{k=1}^K \pi_k p(\mathbf{x} | \mu_k), \quad (1)$$

где $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K\}$, $\boldsymbol{\pi} = \{\pi_1, \dots, \pi_K\}$.

Пусть задана обучающая выборка $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. Для каждого \mathbf{x} введем скрытую переменную $\mathbf{z} = (z_1, \dots, z_K)^T$ — бинарный вектор, у которого только одна компонента равна 1, а все остальные равны 0. Тогда можно записать условное распределение на \mathbf{x} при известной \mathbf{z} :

$$p(\mathbf{x} | \mathbf{Z}, \boldsymbol{\mu}) = \prod_{k=1}^K p(\mathbf{x} | \boldsymbol{\mu}_k)^{z_k}$$

Введём распределение на \mathbf{z} :

$$p(\mathbf{z} | \boldsymbol{\pi}) = \prod_{k=1}^K \pi_k^{z_k}$$

Также введем априорные распределения на параметры $\boldsymbol{\mu}$ и $\boldsymbol{\pi}$:

$$p(\boldsymbol{\pi} | \boldsymbol{\alpha}) = \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \pi_k^{\alpha_k - 1} = \{\boldsymbol{\alpha} = (\alpha, \dots, \alpha)^T\} = \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \prod_{k=1}^K \pi_k^{\alpha - 1}$$

$$p(\boldsymbol{\mu}_k | a, b) = \prod_{i=1}^D \text{Beta}(\mu_{ki} | a, b) = \frac{\mu_{ki}^{a-1} (1 - \mu_{ki})^{b-1}}{B(a, b)}$$

Совместное распределение модели:

$$p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\pi} | \alpha, a, b) = p(\mathbf{X} | \mathbf{Z}, \boldsymbol{\mu}) p(\mathbf{Z} | \boldsymbol{\pi}) p(\boldsymbol{\pi} | \alpha) \prod_{k=1}^K p(\boldsymbol{\mu}_k | a, b)$$

$$p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\pi} | \alpha, a, b) = \prod_{n=1}^N \prod_{k=1}^K \left(\pi_{nk} \prod_{i=1}^D \mu_{ki}^{x_{ni}} (1 - \mu_{ki})^{1-x_{ni}} \right)^{z_{nk}} \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}) \prod_{k=1}^K \prod_{i=1}^D \text{Beta}(\mu_{ki} | a, b) \quad (2)$$

2.2 Е-шаг. Формулы для пересчета $q(\boldsymbol{\mu}), q(\mathbf{Z})$

Запишем логарифм совместного распределения:

$$\begin{aligned} \ln p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\pi} | \alpha, a, b) &= \\ &= \sum_{n=1}^N \sum_{k=1}^K z_{nk} \left[\ln \pi_{nk} + \sum_{i=1}^D \left(x_{ni} \ln \mu_{ki} + (1 - x_{ni}) \ln(1 - \mu_{ki}) \right) \right] + \ln \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}) + \sum_{k=1}^K \sum_{i=1}^D \ln \text{Beta}(\mu_{ki} | a, b) = \\ &= \sum_{n=1}^N \sum_{k=1}^K z_{nk} \left[\ln \pi_{nk} + \sum_{i=1}^D \left(x_{ni} \ln \mu_{ki} + (1 - x_{ni}) \ln(1 - \mu_{ki}) \right) \right] + \ln \Gamma(K\alpha) - K \ln \Gamma(\alpha) + \sum_{k=1}^K (\alpha - 1) \ln \pi_k + \\ &+ \sum_{k=1}^K \sum_{i=1}^D \left((a - 1) \ln \mu_{ki} + (b - 1) \ln(1 - \mu_{ki}) - \ln B(a, b) \right) \end{aligned}$$

На Е-шаге вычисляется вариационное приближение:

$$p(\mathbf{Z}, \boldsymbol{\mu} | \mathbf{X}, \boldsymbol{\pi}, \alpha, a, b) \approx q(\mathbf{Z})q(\boldsymbol{\mu}),$$

Найдём $q(\boldsymbol{\mu})$:

$$\begin{aligned}
 \ln q(\boldsymbol{\mu}) &= \mathbb{E}_{q(\mathbf{Z})} \ln p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\pi} \mid \boldsymbol{\alpha}, a, b) = \\
 &= \mathbb{E}_{q(\mathbf{Z})} \sum_{n=1}^N \sum_{k=1}^K z_{nk} \sum_{i=1}^D \left(x_{ni} \ln \mu_{ki} + (1 - x_{ni}) \ln(1 - \mu_{ki}) \right) + \sum_{k=1}^K \sum_{i=1}^D \left((a-1) \ln \mu_{ki} + (b-1) \ln(1 - \mu_{ki}) \right) + \text{const} = \\
 &= \{ \mathbb{E}[z_{nk}] = r_{nk} \} = \sum_{k=1}^K \sum_{i=1}^D \left(\sum_{n=1}^N \left[r_{nk} (x_{ni} \ln \mu_{ki} + (1 - x_{ni}) \ln(1 - \mu_{ki})) \right] + (a-1) \ln \mu_{ki} + (b-1) \ln(1 - \mu_{ki}) \right) + \text{const} = \\
 &= \sum_{k=1}^K \sum_{i=1}^D \left(\left[\sum_{n=1}^N r_{nk} x_{ni} + a - 1 \right] \ln \mu_{ki} + \left[\sum_{n=1}^N r_{nk} (1 - x_{ni}) + b - 1 \right] \ln(1 - \mu_{ki}) \right) + \text{const}
 \end{aligned}$$

$$q(\boldsymbol{\mu}) = \prod_{k=1}^K \prod_{i=1}^D \text{Beta}(\mu_{ki} \mid \hat{a}_{ki}, \hat{b}_{ki}), \quad \hat{a}_{ki} = a + \sum_{n=1}^N r_{nk} x_{ni}, \quad \hat{b}_{ki} = b + \sum_{n=1}^N r_{nk} (1 - x_{ni}),$$

Запишем формулы для достаточных статистик:

$$\mathbb{E}_{q(\boldsymbol{\mu})} \ln q(\boldsymbol{\mu}) = \psi(\hat{a}) - \psi(\hat{a} + \hat{b}), \quad \mathbb{E}_{q(\boldsymbol{\mu})} \ln(1 - q(\boldsymbol{\mu})) = \psi(\hat{b}) - \psi(\hat{a} + \hat{b}).$$

Формула для пересчёта $\ln q(\mathbf{Z})$:

$$\begin{aligned}
 \ln q(\mathbf{Z}) &= \mathbb{E}_{q(\boldsymbol{\mu})} \ln p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\pi} \mid \boldsymbol{\alpha}, a, b) = \mathbb{E}_{q(\boldsymbol{\mu})} \sum_{n=1}^N \sum_{k=1}^K z_{nk} \left[\ln \pi_{nk} + \sum_{i=1}^D \left(x_{ni} \ln \mu_{ki} + (1 - x_{ni}) \ln(1 - \mu_{ki}) \right) \right] + \text{const} = \\
 &= \sum_{n=1}^N \sum_{k=1}^K z_{nk} \underbrace{\left[\ln \pi_{nk} + \sum_{i=1}^D \left(x_{ni} \mathbb{E}_{q(\boldsymbol{\mu})} \ln \mu_{ki} + (1 - x_{ni}) \mathbb{E}_{q(\boldsymbol{\mu})} \ln(1 - \mu_{ki}) \right) \right]}_{\rho_{nk}} + \text{const}
 \end{aligned}$$

$$\begin{aligned}
 q(\mathbf{Z}) &= \prod_{n=1}^N \prod_{k=1}^K r_{nk}^{z_{nk}} \\
 r_{nk} &= \frac{\exp(\rho_{nk})}{\sum_{j=1}^K \exp(\rho_{nj})}
 \end{aligned}$$

Заметим, что $\mathbb{E}[z_{nk}] = r_{nk}$.

2.3 М-шаг. Формула для пересчёта π

На М-шаге вычисляется точечная оценка на π :

$$\mathbb{E}_{q(\mathbf{Z})q(\boldsymbol{\mu})} \ln p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\pi} \mid \boldsymbol{\alpha}, a, b) \rightarrow \max_{\boldsymbol{\pi}}$$

Запишем Лагранжиан, учитывая, что вектор $\boldsymbol{\pi}$ — вектор вероятностей ($\sum_{k=1}^K \pi_k = 1$).

$$\begin{aligned}
 \mathbf{L} &= \sum_{n=1}^N \sum_{k=1}^K r_{nk} \left[\ln \pi_k + \sum_{i=1}^D \left(x_{ni} \mathbb{E}_{q(\boldsymbol{\mu})} \ln \mu_{ki} + (1 - x_{ni}) \mathbb{E}_{q(\boldsymbol{\mu})} \ln(1 - \mu_{ki}) \right) \right] + \ln \Gamma(K\alpha) - K \ln \Gamma(\alpha) + \\
 &+ \sum_{k=1}^K (\alpha - 1) \ln \pi_k + \sum_{k=1}^K \sum_{i=1}^D \left[(a-1) \mathbb{E}_{q(\boldsymbol{\mu})} \ln \mu_{ki} + (b-1) \mathbb{E}_{q(\boldsymbol{\mu})} \ln(1 - \mu_{ki}) - \ln B(a, b) \right] + \lambda \left(\sum_{k=1}^K \pi_k - 1 \right)
 \end{aligned}$$

$$\begin{aligned}
\frac{d\mathbf{L}}{d\pi_k} &= \sum_{n=1}^N r_{nk} \frac{1}{\pi_k} + (\alpha - 1) \frac{1}{\pi_k} + \lambda = 0, & \longrightarrow & \pi_k = \frac{\sum_{n=1}^N r_{nk} + (\alpha - 1)}{-\lambda}, \\
-\lambda \pi_k &= \sum_{n=1}^N r_{nk} + (\alpha - 1), & \longrightarrow & -\sum_{k=1}^K \lambda \pi_k = \sum_{k=1}^K \left(\sum_{n=1}^N r_{nk} + (\alpha - 1) \right), \\
-\lambda &= N + K(\alpha - 1), & \longrightarrow & \pi_k = \frac{\sum_{n=1}^N r_{nk} + (\alpha - 1)}{N + K(\alpha - 1)}
\end{aligned}$$

2.4 Формула для подсчета вариационной нижней границы

Выпишем формулу для вариационной нижней границы в общем виде:

$$\begin{aligned}
\mathcal{L} &= \int q(\mathbf{Z}, \boldsymbol{\mu}) \ln \frac{p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\pi})}{q(\mathbf{Z}, \boldsymbol{\mu})} = \mathbb{E}_{q(\mathbf{Z})q(\boldsymbol{\mu})} [\ln p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\pi})] - \mathbb{E} [\ln p(\mathbf{Z}, \boldsymbol{\mu})] = \{p(\mathbf{Z}, \boldsymbol{\mu}) \approx q(\mathbf{Z})q(\boldsymbol{\mu})\} = \\
&= \mathbb{E}_{q(\mathbf{Z})q(\boldsymbol{\mu})} [\ln p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\pi})] - \mathbb{E}_{q(\mathbf{Z})} [\ln q(\mathbf{Z})] - \mathbb{E}_{q(\boldsymbol{\mu})} [\ln q(\boldsymbol{\mu})]
\end{aligned}$$

$$\begin{aligned}
\mathcal{L} &= \\
&= \sum_{n=1}^N \sum_{k=1}^K r_{nk} \left[\ln \pi_{nk} + \sum_{i=1}^D \left(x_{ni} \mathbb{E}_{q(\boldsymbol{\mu})} \ln \mu_{ki} + (1 - x_{ni}) \mathbb{E}_{q(\boldsymbol{\mu})} \ln (1 - \mu_{ki}) \right) \right] + \ln \Gamma(K\alpha) - K \ln \Gamma(\alpha) + \sum_{k=1}^K (\alpha - 1) \ln \pi_k + \\
&+ \sum_{k=1}^K \sum_{i=1}^D \left((a - 1) \mathbb{E}_{q(\boldsymbol{\mu})} \ln \mu_{ki} + (b - 1) \mathbb{E}_{q(\boldsymbol{\mu})} \ln (1 - \mu_{ki}) - \ln B(a, b) \right) - \\
&- \sum_{n=1}^N \sum_{k=1}^K r_{nk} \ln r_{nk} - \sum_{k=1}^K \sum_{i=1}^D \left((\hat{a}_{ki} - 1) \mathbb{E}_{q(\boldsymbol{\mu})} \ln \mu_{ki} + (\hat{b}_{ki} - 1) \mathbb{E}_{q(\boldsymbol{\mu})} \ln (1 - \mu_{ki}) - \ln B(\hat{a}_{ki}, \hat{b}_{ki}) \right)
\end{aligned}$$

3 Модельные данные

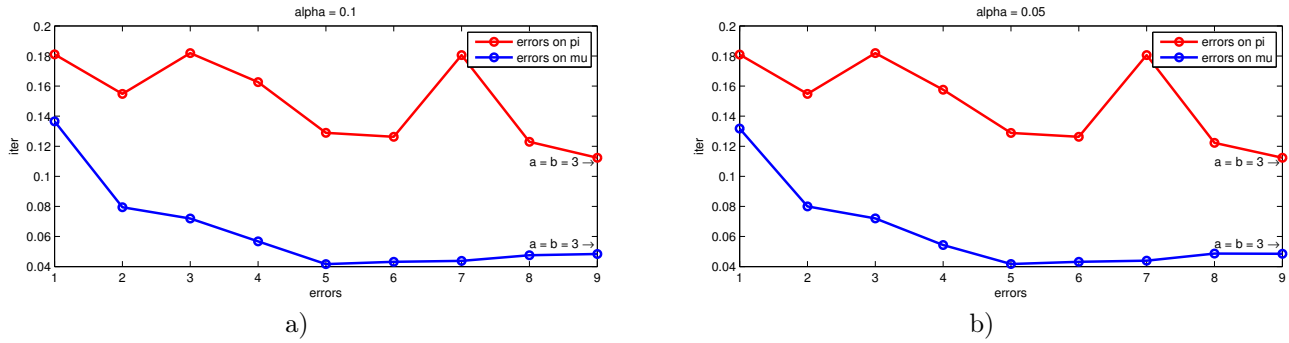
Протестируем полученный ЕМ-алгоритм на модельных данных. Сгенерируем выборку из заданного распределения (2) с параметрами $a = b = 3$, $\alpha = 0.01$, $K = 70$. ЕМ-алгоритм будем запускать на сетке значений

$$\mathbf{a} = [1, 2, 3], \quad \mathbf{b} = [1, 2, 3], \quad \boldsymbol{\alpha} = [0.01, 0.5, 0.1].$$

Для того, чтобы оценить качество работы ЕМ-алгоритма сравним априорные значения $\boldsymbol{\pi}, \boldsymbol{\mu}$ с значениями, которые выдает ЕМ-алгоритм ($\boldsymbol{\pi}_{EM}, \boldsymbol{\mu}_{EM}$). Для сравнения будем использовать следующую метрику: отсортируем вектора $\boldsymbol{\pi}, \boldsymbol{\pi}_{EM}$ по возрастанию и посчитаем модуль разности между ними. Отсортируем $\boldsymbol{\mu}, \boldsymbol{\mu}_{EM}$ по возрастанию $\boldsymbol{\pi}, \boldsymbol{\pi}_{EM}$ соответственно и посчитаем модуль среднего отклонения μ_{ki} от $(\boldsymbol{\mu}_{EM})_{ki}$. На рис. 1 показаны графики зависимости ошибки на $\boldsymbol{\pi}, \boldsymbol{\mu}$, в зависимости от a, b . Видно, что в целом ошибка небольшая, и минимум ошибки $\boldsymbol{\pi}$ достигается при $a = b = 3$, значит ЕМ-алгоритм правильно восстановил данные.

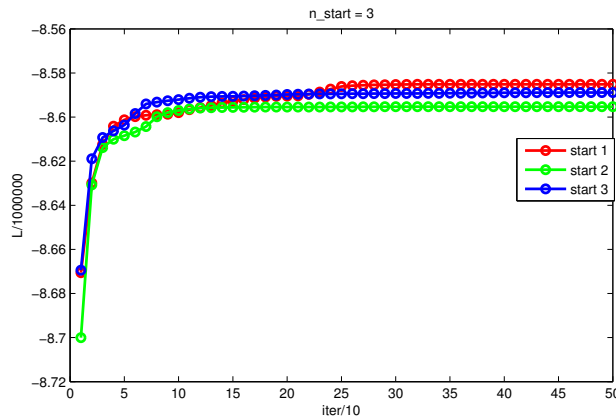
4 Данные MNIST

Протестируем полученный алгоритм на базе MNIST. Подготовленные бинаризованные данные MNIST были скачаны по ссылке <https://yadi.sk/d/SSdrHQtWcZiEY>. Это csv файл с 60000 строк, в каждой строке записано бинарное изображение цифры от 0 до 9 и метка класса. Каждое изображение имеет размер 28 на 28 пикселей, соответственно в строке записано 785 чисел. (i, j) пиксель изображения записан в позиции 28i + j, нумерация с нуля. Последнее число в строке — метка класса.

Рис. 1. а) $\alpha = 0.01$, б) $\alpha = 0.05$.

4.1 Функционал $\mathcal{L}(q)$

Убедимся, что с увеличением количества итераций функционал увеличивается. Запустим ЕМ-алгоритм с числом запусков из различных случайных начальных приближений равным 3 (см. рис. 2). Действительно, функционал возрастает, при различных начальных значениях получаются различные конечные значения функционала.

Рис. 2. Зависимость $\mathcal{L}(q)$ от числа итераций

4.2 Зависимость числа получаемых кластеров от априорного значения количества компонент

Посмотрим зависимость K (числа получаемых кластеров) от априорного значения количества компонент, подаваемого в ЕМ-алгоритм (`options.K`) при фиксированных параметрах α, a, b (см. таблицу 1). Из таблицы видно, что при `options.K` ≤ 50 полученное число кластеров примерно равно `options.K`. Это обусловлено тем, что при малых значениях априорного числа компонент ЕМ-алгоритм выделяет столько же кластеров, сколько и подавалось в алгоритм, т.к. в данных можно выделить достаточно много кластеров и никакие веса не зануляются. При `options.K` > 50 выделяется примерно одинаковое число кластеров (50 - 75), ЕМ-алгоритм сам сокращает априорное число количества компонент до нужного оптимального значения.

	$a = b = 0.5$ $\alpha = 0.001$
<code>options.K</code> = 5	5
<code>options.K</code> = 10	10
<code>options.K</code> = 50	49
<code>options.K</code> = 100	58
<code>options.K</code> = 300	71
<code>options.K</code> = 400	80
<code>options.K</code> = 500	75

Таблица 1. Зависимость количества получаемых кластеров от начального значения K .

4.3 Зависимость числа получаемых кластеров от α

Посмотрим зависимость K (числа получаемых кластеров) от α (см. рис. 3). Все графики при различных значениях α очень похожи, при малых значениях $a = b$ графики сливаются в один. Можно сделать вывод, что от параметра α количество выделенных кластеров не зависит.

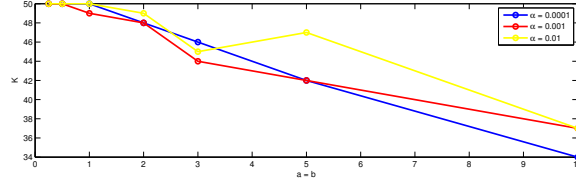


Рис. 3. Зависимость K от α , $a = b$.

4.4 Зависимость числа получаемых кластеров от a, b

При $a = b$ (см. рис. 4) можно заметить, что с увеличением a, b количество выделяемых кластеров уменьшается. При фиксированном $b = 1$, при малых значениях a в экспериментах выделялось 49-50 кластеров, это было связано с тем, что данным при малых a, b "не хватало" кластеров; при $\text{options.K} = 500$ (см. рис. 5) количество выделяемых кластеров при увеличении a тоже уменьшается.

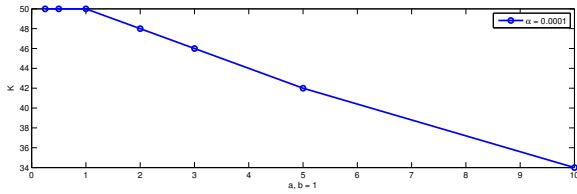


Рис. 4. а) Зависимость K от $a = b$.

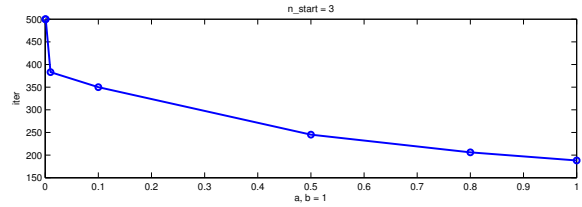


Рис. 5. б) Зависимость K от $a, b = 1$.

4.5 Визуализация полученных центров кластеров $\mathbb{E}_{q(\mu)}\mu$

Визуализируем центры кластеров, полученных при различных параметрах алгоритма (см. рис 4.5). На рисунке а) алгоритм выделил 48 кластеров, цифры чёткие. Можно заметить, что все начертания цифры нуль (и остальных цифр соответственно) отличаются друг от друга (поэтому много кластеров). На рисунке б) алгоритм выделил 6 кластеров, цифры расплывчатые, то есть алгоритм относит разные цифры к одному кластеру: описывает одним кластером много объектов — из-за этого расплывчатость.

5 Исследование логарифма правдоподобия

Исследуем зависимость логарифма правдоподобия на обучающей и контрольной выборках от кластеризации. Правдоподобие вычисляется по формуле

$$p(\mathbf{X}) = \prod_{n=1}^N p(\mathbf{x}_n | \mathbb{E}_{q(\mu)}\mu, \pi_{ML}),$$

где $p(\mathbf{x} | \mu, \pi)$ задано формулой 1, π_{ML} — точечная оценка на параметры π , полученная в результате работы ЕМ-алгоритма.

$$p(\mathbf{X}) = \prod_{n=1}^N \sum_{k=1}^K (\pi_{ML})_k \left(\prod_{i=1}^D \mathbb{E}_{q(\mu)}\mu_{ki}^{x_{ni}} (1 - \mathbb{E}_{q(\mu)}\mu_{ki})^{1-x_{ni}} \right)$$

$$\ln p(\mathbf{X}) = \sum_{n=1}^N \ln \sum_{k=1}^K (\pi_{ML})_k \exp \left(\sum_{i=1}^D x_{ni} \ln \mathbb{E}_{q(\mu)}\mu_{ki} + (1 - x_{ni}) \ln(1 - \mathbb{E}_{q(\mu)}\mu_{ki}) \right)$$

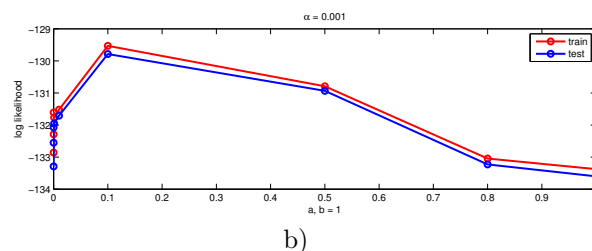
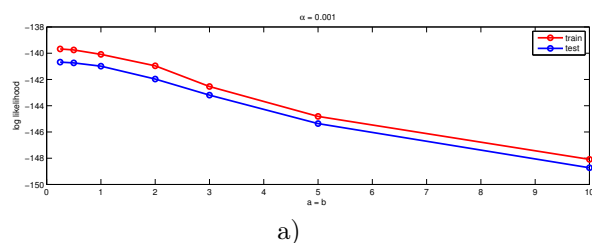


Рис. 6. а) количество выявленных кластеров = 48



Рис. 7. б) количество выявленных кластеров = 5

Графики логарифма правдоподобия для тестовой и обучающей выборки очень похожи. На графике б) 8 при малых значениях a происходит скачок вниз, он объясняется тем, что выделяется слишком много кластеров.

Рис. 8. а) $a = b$, б) $a, b = 1$.

6 Классификация с новым признаком для n -го объекта: величины $q(z_{nk} = 1)$. Матрица точности.

Рассмотрим величины $q(z_{nk} = 1)$ в качестве признаков n -го объекта. В качестве классификатора будем использовать svm из пакета libsvm. При различных параметрах a, b, α точность варьируется в районе 80%, что является хорошим результатом (с учетом того, что параметры c, γ были установлены по умолчанию). Рассмотрим матрицу точности на контрольной выборке при $a = b = 1$ (см. таблицу 6). Все остальные параметры были установлены по умолчанию. Самое плохое качество — при классификации цифр 4 и 9, они часто "путаются" друг с другом. Также цифра 5 часто принимается за тройку, всё это объяснимо их похожими написаниями. При $a = 0.01, b = 1$ ситуация практически ничем не отличается. При $a = b = 10$ (см. таблицу 6) ЕМ-алгоритмом было выделено меньше кластеров, поэтому значительно увеличилось ошибки при классификации цифры 9 (она стала также ложно относиться ещё и к классу 7). Также увеличился процент троек, неверно классифицируемых как восьмерки. Данный результат был ожидаем, т.к. при уменьшении числа выделяемых кластеров больше цифр относятся к одному кластеру.

Как уже было сказано ранее точность классификации при различных параметрах составила примерно 80%, учитывая, что ЕМ-алгоритм значительно сократил размерность признакового пространства (с 784 признаков до 50), можно считать, что данный метод классификации является хорошим.

	0	1	2	3	4	5	6	7	8	9
0	94.1473	0	0.5045	0.5045	0.1009	1.3118	1.1100	0	2.3209	0
1	0	95.8647	0.4699	0.6579	0.0940	0.3759	0.5639	0	1.6917	0.2820
2	0.5051	0.1010	93.9394	0.8081	0.5051	0.2020	0.1010	0.2020	3.6364	0
3	0.3883	0.1942	1.1650	80.7767	0.1942	7.3786	0	0	8.4466	1.4563
4	0.1017	0.5086	0.3052	0.2035	72.0244	0.3052	0.8138	1.6277	0.4069	23.7030
5	0.6557	0	0.4372	15.8470	1.3115	70.9290	2.8415	0.1093	6.6667	1.2022
6	0.7239	0.1034	0	0	0.2068	0.6205	97.8283	0	0.5171	0
7	0.0917	0.7339	1.7431	0.3670	2.8440	0.2752	0	80.8257	0.7339	12.3853
8	0.4955	0.4955	0.3964	3.8652	0.4955	2.6759	0.2973	0.3964	90.2874	0.5946
9	0.5203	0	0.1041	1.6649	22.3725	0.2081	0	13.2154	1.7690	60.1457

Таблица 2. $a = b = 1$, $\alpha = 0.001$, accuracy = 83.88 %

	0	1	2	3	4	5	6	7	8	9
0	91.1201	0	2.3209	0.4036	0	1.2109	1.5136	0	3.4309	0
1	0	95.3008	0.7519	0.5639	0.0940	0.1880	0.6579	0.1880	2.2556	0
2	0.4040	0	94.8485	1.0101	0	0	0.6061	0	3.1313	0
3	0.1942	0	1.9417	80.0000	0.0971	9.2233	0	0.4854	7.2816	0.7767
4	0	0.4069	4.2726	0	65.2085	1.5259	0.5086	4.6796	0.4069	22.9908
5	0.5464	0	0.9836	15.7377	1.2022	69.2896	2.8415	0.4372	8.0874	0.8743
6	0.9307	0	2.0683	0	0	0.6205	95.9669	0	0.4137	0
7	0.0917	0.5505	2.2936	0.2752	4.2202	0.1835	0.0917	75.9633	0.7339	15.5963
8	0.5946	0.2973	1.8831	5.5500	0.0991	2.7750	0.2973	0.1982	86.9177	1.3875
9	0.3122	0.1041	0.4162	1.6649	20.3954	0.2081	0	11.2383	2.1852	63.4755

Таблица 3. $a = 0.01$, $b = 1$, $\alpha = 0.001$, accuracy = 81.98 %

	0	1	2	3	4	5	6	7	8	9
0	92.7346	0	0.6054	0.1009	0.1009	1.4127	2.2200	0	2.8254	0
1	0	95.9586	0.2820	0.0940	0	0.6579	0.4699	0.0940	2.3496	0.0940
2	0.6061	0.3030	91.5152	0.1010	0.4040	0.2020	0.3030	0.3030	6.2626	0
3	0.1942	0.1942	1.1650	61.8447	0.0971	9.1262	0.0971	0.0971	26.2136	0.9709
4	0.1017	0.5086	0.9156	0	55.9512	1.1190	0.8138	3.7640	0.9156	35.9105
5	1.0929	0	0.3279	18.6885	1.3115	61.4208	2.7322	0.3279	13.0055	1.0929
6	0.6205	0.3102	0.1034	0	0	0.9307	97.4147	0	0.6205	0
7	0	0.9174	1.3761	0	2.2936	0.2752	0	79.6330	1.1009	14.4037
8	0.5946	0.6938	0.6938	3.7661	0.4955	3.2706	0.1982	1.4866	88.3053	0.4955
9	0.4162	0.3122	0.1041	1.2487	16.2331	0.3122	0	16.8574	1.8730	62.6431

Таблица 4. $a = b = 10$, $\alpha = 0.001$, accuracy = 78.98 %

7 Выводы

Работа ЕМ-алгоритма зависит от подаваемых параметров, однако в разной степени, например, от параметра α зависит не сильно, а от a, b зависимость значительнее.

ЕМ-алгоритм оказался хорошим методом сокращения размерности.