ЛЕКЦИЯ 7. АНАЛИЗ ГЛАВНЫХ КОМПОНЕНТ

Сергей Николенко

1. Размерности

Пусть заданы два пространства: $\mathbb{R}^n$ и $\mathbb{R}^m \subset \mathbb{R}^n,\ m < n$. При переходе в какой-либо задаче к пространству меньшей размерности необходимо спроектировать в него большее пространство таким образом, чтобы не испортились важные свойства. Например, рассмотрим равномерное распределение точек на сфере в $\mathbb{R}^n$. Выделим какую-либо прямую, проходящую через сферу. Спроектируем на нее точки сферы. В каждой точке прямой плотность распределения определяется формулой:

$$\rho(\xi) \sim (1-\xi^2)^{\frac{n-1}{2}},$$

где n —размерность пространства, $\xi = \cos\theta$.

В общем случае полученное распределение близко к гауссовому, а в $\mathbb{R}^3$ совпадает с распределением для V_1.

2. Приближение векторов

Пусть задан $\bar{x} = \{\vec{x}_i|\ i = 1..m\}$ — набор векторов. Необходимо найти $\vec{n}$ — вектор, вдоль которого лучше всего описывается $\bar{x}$. Предположим сначала, что все $\vec{x}_i$ лежат на одной прямой, то есть :

$$\vec{x}_i = \vec{\mu} + \theta_i \cdot \vec{n},\ |\vec{n}| = 1$$

Сосчитаем следующую величину:

$$S_n = \frac{1}{m}\sum_i^m \left((\vec{x}_i - \vec{\mu})\cdot\vec{n}\right)^2 = \frac{1}{m}\sum_i^m {\theta_i}^2$$

Пусть теперь за наилучшее описание выбран какой-либо другой вектор $\vec{n}'$:

$$S_{n'} = \frac{1}{m}\sum_i^m \left((\vec{x}_i - \vec{\mu})\cdot\vec{n}'\right)^2 = \frac{1}{m}\sum_i^m {\theta_i}^2(\vec{n}\cdot\vec{n}')^2$$

Тогда приходим к идее, что лучше всего подходит такой вектор $\vec{n}'$, который *максимизирует* значение $S_{n'}$.

3. Метод анализа главных компонент

Запишем *выборочную дисперсию* вдоль вектора $\vec{n}$:

$$\vec{n}^T \cdot C \cdot \vec{n},$$

где C —*матрица ковариаций*, то есть:

$$C_{ij} = (\vec{x}_i - \vec{\mu})\cdot(\vec{x}_j - \vec{\mu})$$

Законспектировали Сергей Гиндин и Андрей Давыдов.

Отметим, что C является матрицей Грамма для векторов $(\vec{x}_i - \vec{\mu})$ и, следовательно, она симметрична.

C —положительная полуопределенная матрица (Доказательство оставим читателю в качестве упражнения):

$$\forall \vec{z} : \vec{z}^T \cdot C \cdot \vec{z} > 0,$$

поэтому, если у C есть собственный вектор $(C \cdot \ \vec{e} = \lambda \cdot \vec{e})$, то

$$\vec{e}^T \cdot C \cdot \vec{e} = \lambda(\vec{e}^T \cdot \vec{e}) = \lambda|\vec{e}|^2 > 0 \Rightarrow \lambda > 0$$

Пусть теперь найден ортонормированный набор базисных векторов $\{\vec{e}_1 \dots \vec{e}_d\}$, которому соответствуют собственные числа $\lambda_1 \geq \lambda_2 \ \geq \cdots \geq \lambda_d > 0$. Тогда можно представить $\vec{n} = \sum_a^d \alpha_a \vec{e}_a$ и переписать:

$$\vec{n}^T \cdot C \cdot \vec{n} = \left(\sum_a^d \alpha_a \vec{e}_a\right)^T \cdot C \cdot \left(\sum_a^d \alpha_a \vec{e}_a\right) = \left(\sum_a^d \alpha_a \vec{e}_a\right)^T \cdot \left(\sum_a^d (\lambda_a \alpha_a)\vec{e}_a\right) = \sum_a^d \lambda_a {\alpha_a}^2$$

Необходимо максимизировать полученное выражение: $\sum_a^d \lambda_a {\alpha_a}^2$ с учетом условия $\sum {\alpha_a}^2 = 1$.

Выражение будет максимальным, если $|\alpha_1| = 1$, то есть нужно принять $\vec{n} = \pm \vec{e}_1$. Теперь, чтобы найти следующую компоненту, нужно спроектировать все в подпространство $\perp \vec{e}_1$, то есть положить $\alpha_1 = 0$.

Этот метод получил название PCA (Principal Component Analysis) —анализ главных компонентов.

Теперь рассмотрим другой базис $\vec{u}_a, \ a = 1..d$:

$$\forall \vec{z} = \sum_a^d z_a \vec{u}_a \Rightarrow \vec{z} = \sum_a^d \tilde{z}_a \vec{e}_a, \text{ где } \tilde{z}_a = \vec{z} \cdot \vec{e}_a$$

В этом случае марица C будет выглядеть так (доказательство предлагается в качестве упражнения):

$$C = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_d \end{pmatrix}$$

4. Некоторые свойства

Рассмотрим проекцию вектора $\vec{x}$ на d' первых компонентов в его разложении.

$$\vec{x}' = \sum_a^{d'} \tilde{x}_a e_a$$

Перечислим некоторые интересующие нас свойства метода PCA.

(1) *Декореллирование* — PCA декореллирует компоненты векторных данных.

(2) *Минимизация ошибки* — при проецировании накопилась ошибка:

$$\sum_i^m \|\vec{x}_i - \vec{x}'_i\|^2$$

PCA минимизирует эту ошибку.

Доказательство. Действительно: выберем произвольный ортонормированный базис:

$$\vec{x} = \sum_a^d x_a \vec{u}_a$$

$$\vec{x}' = \sum_a^{d'} x_a \vec{u}_a$$

Преобразуем квадрат нормы разности:

$$\|\vec{x}_i - \vec{x}_i'\|^2 = \|x_i\|^2 - 2(\vec{x}_i \cdot \vec{x}_i') + \|\vec{x}_i'\|^2 = \|x_i\|^2 - 2\sum_a^{d'} x_a^2 + \sum_a^{d'} x_a^2 = \|x_i\|^2 - \sum_a^{d'} (\vec{x} \cdot \vec{u}_a)^2$$

Тогда:

$$\frac{1}{m}\sum_i^m \|\vec{x}_i - \vec{x}_i'\|^2 = \frac{1}{m}\sum_i^m \|\vec{x}_i\|^2 - \frac{1}{m}\sum_a^{d'}\sum_i^m (\vec{x}_i \cdot \vec{u}_a)^2$$

Заметим, что

$$\sum_i^m (\vec{x}_i \cdot \vec{u}_a)^2 = \sum_i^m \left((\vec{u}_a^T \cdot \vec{x}_i)(\vec{x}_i^T \cdot \vec{u}_a)^T\right) = \sum_i^m (\vec{u}_a^T \cdot \vec{x}_i) \sum_i^m (\vec{x}_i^T \cdot \vec{u}_a) =$$

$$= \vec{u}_a^T \left(\left(\sum \vec{x}_i\right) \cdot \left(\sum \vec{x}_i\right)^T\right) \vec{u}_a = \vec{u}_a^T \cdot C \cdot \vec{u}_a$$

Разложив $\vec{u}_a$ по исходному базису: $\vec{u}_a = \sum_a^d \beta_{ap}\vec{e}_p$, где $\sum_p^d \beta_{ap}^2 = 1$, получаем оценку ошибки:

$$\frac{1}{m}\left(\sum_i^m \|\vec{x}_i\|^2 - \sum_a^{d'}\sum_i^m (\vec{x}_i \cdot \vec{u}_a)^2\right) =$$

$$= \frac{1}{m}\left(\sum_i^m \|\vec{x}_i\|^2 - \sum_a^{d'} (\vec{u}_a^T \cdot C \cdot \vec{u}_a)\right) =$$

$$= \frac{1}{m}\left(\sum_i^m \|\vec{x}_i\|^2 - \sum_a^{d'}\sum_p^d (\beta_{ap}^2 \lambda_p)\right)$$

Чтобы минимизировать ошибку, нужно максимизировать выражение:

$$\sum_a^{d'}\sum_p^d (\beta_{ap}^2 \lambda_p).$$

Нетрудно показать, что этому условию соответствует выбор $\beta_{ij} = \delta_{ij}$, где

$$\delta_{ij} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$$

$\square$

Итак, мы доказали, что функция ошибки минимизируется если выбрать первые d' векторов, совпадающих с e_i.

5. Практическое применение

Пусть есть набор векторов $X = (\vec{x}_1 \dots \vec{x}_m) \in \mathbb{R}^d$

(1) Находим $\vec{\mu} = \frac{1}{m} \sum_i^m \vec{x}_i$
(2) Усредняем набор $B = (\vec{x}_1 - \vec{\mu} \dots \vec{x}_m - \vec{\mu})$
(3) Считаем матрицу ковариаций: $C = \frac{1}{m}(B \cdot B^T)$
(4) Ищем собственные векторы и числа для C
(5) Получаем диагонализированную матрицу V: $V^{-1} \cdot C \cdot V$
(6) Оставляем только первые d' компонент $V_{d\times d} \longrightarrow W_{d'\times d}$
(7) Считаем нормализованную матрицу $Z = B/\sqrt{C_i}$
(8) Получаем проекцию $Y = W^T \cdot Z$

Последующие разделы конспекта (вторая часть) посвящена применению обучения по Хеббу для анализа главных компонент.

6. Применение фильтра Хебба для выделения максимального собственного значения

Покажем, как линейный нейрон с Хеббовским правилом адаптации синаптических весов может использоваться в качестве фильтра для выделения первого главного компонента вектора входных сигналов.

Пусть выходной сигнал нейрона равен линейной комбинации входных сигналов.

$$y = \sum_{i=1}^{m} w_i x_i \tag{1}$$

Наша задача подобрать веса w_i таким образом, чтобы y был равен максимальному главному компоненту вектора $\mathbf{x} = \{x_i\}$.

Согласно постулату Хебба, синаптический вес w_i возрастает, если предсинаптический сигнал x_i и постсинаптический сигнал y совпадают друг с другом. Это правило можно выразить, к примеру, следующим образом:

$$w_i(n+1) = w_i(n) + \eta y(n) x_i(n), i = 1, 2, \dots, m,$$

где n — номер шага по времени; η — *параметр интенсивности обучения.*

Проблема в том, что это правило обучения приводит к неограниченному росту синаптических весов. Это можно обойти используя нормализацию:

$$w_i(n+1) = \frac{w_i(n) + \eta y(n) x_i(n)}{(\sum_{j=1}^{m} [w_j(n) + \eta y(n) x_j(n)]^2)^{1/2}} \tag{2}$$

Заметим, что Евклидова норма определенного таким образом вектора $\mathbf{w} = \{w_i\}$ равна 1:

$$\|\mathbf{w}(n+1)\|^2 = \sum_{i=1}^{m} w_i^2(n+1) = \sum_{i=1}^{m} \frac{[w_i(n) + \eta y(n) x_i(n)]^2}{\sum_{j=1}^{m} [w_j(n) + \eta y(n) x_j(n)]^2} =$$
$$= \frac{\sum_{i=1}^{m} [w_i(n) + \eta y(n) x_i(n)]^2}{\sum_{j=1}^{m} [w_j(n) + \eta y(n) x_j(n)]^2} = 1$$

Использование нормализации обеспечивает введение конкуренции между синапсами нейрона за облдание ограниченными ресурсами, которая является существенным условием стабилизации сети (пока что состоящей из одного нейрона). Предположим, что η мало, тогда

$$w_i(n+1) = w_i(n) + \eta y(n)[x_i(n) - y(n)w_i(n)] + O(\eta^2)$$

Доказательство. Это доказывается несколькими применениями формулы Тейлора:

$$w_i(n+1) = \frac{w_i(n) + \eta y(n)x_i(n)}{(\sum_{j=1}^m [w_j(n) + \eta y(n)x_j(n)]^2)^{1/2}} =$$

$$= \frac{w_i(n) + \eta y(n)x_i(n)}{(\sum_{j=1}^m [w_j(n)^2 + 2\eta y(n)w_j(n)x_j(n) + O(\eta^2)])^{1/2}} =$$

$$= \frac{w_i(n) + \eta y(n)x_i(n)}{(\sum_{j=1}^m [w_j(n)^2] + 2\eta y(n)\sum_{j=1}^m w_j(n)x_j(n) + O(\eta^2))^{1/2}} = \frac{w_i(n) + \eta y(n)x_i(n)}{(1 + 2\eta y^2(n) + O(\eta^2))^{1/2}} =$$

$$= \frac{w_i(n) + \eta y(n)x_i(n)}{1 + \eta y^2(n) + O(\eta^2)} = [w_i(n) + \eta y(n)x_i(n)][1 - \eta y^2(n) + O(\eta^2)] =$$

$$= w_i(n) + \eta[y(n)x_i(n) - w_i(n)y^2(n)] + O(\eta^2) = w_i(n) + \eta y(n)[x_i(n) - y(n)w_i(n)] + O(\eta^2)$$

□

Таким образом, мы можем аппроксимировать (2) следующим образом:

$$w_i(n+1) = w_i(n) + \eta y(n)[x_i(n) - y(n)w_i(n)]. \tag{3}$$

Слагаемое $y(n)x_i(n)$ участвует в процессе самоуселения. Отрицательное слагаемое $(-y^2(n)w_i(n))$ отвечает за стабилизацию.

7. Матричная формулировка алгоритма

Введем обозначения

$$\mathbf{x}(n) = [x_1(n), x_2(n), \ldots, x_m(n)]^T$$
$$\mathbf{w}(n) = [w_1(n), w_2(n), \ldots, w_m(n)]^T.$$

Тогда формулы (1), (3) преобразуются следующим образом:

$$y(n) = \langle \mathbf{x}(n), \mathbf{w}(n)\rangle = \mathbf{x}^T(n)\mathbf{w}(n) = \mathbf{w}^T(n)\mathbf{x}(n) \tag{4}$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta y(n)[\mathbf{x}(n) - y(n)\mathbf{w}(n)]. \tag{5}$$

Подставляя (4) в (5) получаем:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta[\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{w}(n) - \mathbf{w}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{w}(n)\mathbf{w}(n)]. \tag{6}$$

8. Доказательство корректности фильтра Хебба

Алгоритм обучения (6) является частным случаем алгоритма стохастической аппроксимации:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta(n)h(\mathbf{w}(n), \mathbf{x}(n)), n = 0, 1, 2, \ldots \tag{7}$$

Пусть для величин из алгоритма (7) выполнены следующие условия.

(1) $\eta(n)$ — убывающая последовательность положительных действительных чисел, такая что:

(a)

$$\sum_{n=1}^{\infty} \eta(n) = \infty,$$

(b)

$$\sum_{n=1}^{\infty} \eta^p(n) < \infty, p > 1,$$

(c)

$$\eta(n) \to 0 \text{ при } n \to \infty.$$

(2) Последовательность $\mathbf{w}(n)$ ограничена с вероятностью 1.
(3) *Функция коррекции* $h(\mathbf{w}, \mathbf{x})$ непрерывно дифференцируема по $\mathbf{w}$ и $\mathbf{x}$, и ее производные ограничены по времени.
(4) Предел

(8) $$\bar{h} = \lim_{n\to\infty} E[h(\mathbf{w}, \mathbf{X})]$$

существует для всех $\mathbf{w}$. $\mathbf{X}$ — случайная величина, принимающая значения $\mathbf{x}$.
(5) Существует локально асимпотически устойчивое (в смысле Ляпунова) решение обычного дифференциального уравнения

(9) $$\frac{d}{dt}\mathbf{w}(t) = \bar{h}(\mathbf{w}(t)),$$

где t — непрерывное время.

Устойчивость в смысле Ляпунова. Рассмотрим автономную динамическу систему описываемую уравнением

(10) $$\frac{d}{dt}\mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t)).$$

Определение 1. *Вектор* $\bar{\mathbf{x}}$ *называется стационарным состоянием системы, если* $\mathbf{F}(\bar{\mathbf{x}}) = \mathbf{0}$.

Очевидно, что $\mathbf{x}(t) = \bar{\mathbf{x}}$ будет решением системы (10).

Определение 2. *Стационарное состояние* $\bar{\mathbf{x}}$ *называется равномерно устойчивым, если для любого положительного* ϵ *существует такое положительное* δ, *что из условия*

$$\|\mathbf{x}(0) - \bar{\mathbf{x}}\| < \delta$$

следует, что

$$\|\mathbf{x}(t) - \bar{\mathbf{x}}\| < \epsilon \text{ для всех } t > 0.$$

Грубо говоря, если начальное состояние $\mathbf{x}(0)$ близко к $\bar{\mathbf{x}}$, то траектория системы (10) лежит в малой окрестности $\bar{\mathbf{x}}$.

Определение 3. *Стационарное состояние* $\bar{\mathbf{x}}$ *назывется сходящимся, если существует такое положительное* δ, *что из условия*

$$\|\mathbf{x}(0) - \bar{\mathbf{x}}\| < \delta$$

следует, что

$$\mathbf{x}(t) \to \bar{\mathbf{x}} \text{ при } t \to \infty.$$

Грубо говоря, если начальное состояние $\mathbf{x}(0)$ близко к $\bar{\mathbf{x}}$, то траектория системы (10) достигает $\bar{\mathbf{x}}$ при устремлении t к бесконечности.

Определение 4. *Стационарное состояние $\bar{\mathbf{x}}$ называется асимтотически устойчивым, если оно одновременно является устойчивым и сходящимся.*

Теорема 1. *Стационарное состояние $\bar{\mathbf{x}}$ является устойчивым, если в малой окрестности этой точки существует положительно определенная функция $V(\mathbf{x})$, такая, что ее производная по времени в этой области является отрицательно полуопределенной.*

Теорема 2. *Стационарное состояние $\bar{\mathbf{x}}$ является асимптотически устойчивым, если в малой окрестности этой точки существует положительно определенная функция $V(\mathbf{x})$, такая, что ее производная по времени в этой области является отрицательно определенной.*

Функция $V(\mathbf{x})$, удолетворяющая всем этим требованиям, называется функцией Ляпунова для стационарного состояния $\bar{\mathbf{x}}$. Грубо говоря, это кинетическая энергия системы (10).

Функция $V(\mathbf{x})$ называется положительно определенной, если выполняются следующие условия.

(a) $V(\mathbf{x})$ имеет непрерывные частные производные по всем компонентам вектора $\mathbf{x}$.

(b) $V(\bar{\mathbf{x}}) = 0$.

(c) $V(\mathbf{x}) > 0$ при $\mathbf{x} \neq \bar{\mathbf{x}}$.

Теорема 3. *При выполнении условий (1)–(5)*

$$\lim_{n\to\infty} \mathbf{w}(n) = \mathbf{q}, \text{ бесконечно часто с вероятнотью } 1, \tag{11}$$

где $\mathbf{q}$ — решение уравнения (9).

Приведем уравнение (6) к виду (7). Пусть

$$\eta(n) = \frac{1}{n},$$

условие (1) будет выполнено;

$$h(\mathbf{w}, \mathbf{x}) = \mathbf{x}(n)\mathbf{x}^T(n)\mathbf{w}(n) - [\mathbf{w}^T(n)\mathbf{x}(n)\mathbf{x}^T(n)\mathbf{w}(n)]\mathbf{w}(n), \tag{12}$$

условие (3) также, очевидно, выполнено.

$$\begin{aligned}\bar{h} = \lim_{n\to\infty} E[\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{w}(n) - (\mathbf{w}^T(n)\mathbf{X}(n)\mathbf{X}^T(n)\mathbf{w}(n)]\mathbf{w}(n))] = \\ = \mathbf{R}\mathbf{w}(\infty) - [\mathbf{w}^T(\infty)\mathbf{R}\mathbf{w}(\infty)]\mathbf{w}(\infty),\end{aligned} \tag{13}$$

где $\mathbf{R}$ — матрица корреляции $\mathbf{X}$, $\mathbf{w}(\infty)$ — предельное значение $\mathbf{w}$. Таким образом, условие (4) выполнено.

Для выполнения условия 5 найдем устойчивые точки дифференциального уравнения

$$\frac{d}{dt}\mathbf{w}(t) = \bar{h}(\mathbf{w}(t)) = \mathbf{R}\mathbf{w}(t) - [\mathbf{w}^T(t)\mathbf{R}\mathbf{w}(t)]\mathbf{w}(t). \tag{14}$$

Рассмотрим разложение $\mathbf{w}(t)$ в базисе нормализованных собственных векторов $\{\mathbf{q}_k\}$ матрицы $\mathbf{R}$:

$$\mathbf{w}(t) = \sum_{k=1}^{m} \theta_k(t)\mathbf{q}_k \tag{15}$$

Подставим (15) в (14) и воспользуемся следующими определениями

$$\mathbf{R}\mathbf{q_k} = \lambda_k \mathbf{q}_k$$

и

$$\mathbf{q_k^T}\mathbf{R}\mathbf{q_k} = \lambda_k,$$

где λ_k — собственное число, соответствующее вектору $\mathbf{q}_k$. Получим:

$$\sum_{k=1}^{m} \frac{d\theta_k(t)}{dt}\mathbf{q}_k = \sum_{k=1}^{m} \lambda_k \theta_k(t)\mathbf{q}_k - \left[\sum_{l=1}^{m} \lambda_l \theta_l^2(t)\right] \sum_{k=1}^{m} \theta_k(t)\mathbf{q}_k,$$

или пользуясь линейной независимостью набора $\{\mathbf{q}_k\}$:

$$\frac{d\theta_k(t)}{dt} = \lambda_k \theta_k(t) - \left[\sum_{l=1}^{m} \lambda_l \theta_l^2(t)\right] \theta_k(t). \tag{16}$$

Рассмотрим два случая.

Случай 1. $1 < k \le m$**.** Определим

$$\alpha_k(t) = \frac{\theta_k(t)}{\theta_1(t)}, 1 < k \le m. \tag{17}$$

Предполагается, что $\theta_1(t) \ne 1$ с вероятностью 1. Дифференцируя по t уравнение (17) получим:

$$\frac{d\alpha_k(t)}{dt} = \frac{1}{\theta_1(t)} \frac{d\theta_k(t)}{dt} - \frac{\theta_k(t)}{\theta_1^2(t)} \frac{d\theta_1(t)}{dt} = \frac{1}{\theta_1(t)} \frac{d\theta_k(t)}{dt} - \frac{\alpha_k(t)}{\theta_1(t)} \frac{d\theta_1(t)}{dt}.$$

Подставив (16) и (17) и упростив, находим $\alpha_k(t)$:

$$\frac{d\alpha_k(t)}{dt} = -(\lambda_1 - \lambda_k)\alpha_k(t), 1 < k \le m,$$

то есть

$$\alpha_k(t) = \alpha_k(0) * \exp(-(\lambda_1 - \lambda_k)t).$$

Таким образом

$$\lim_{t\to\infty} \theta_k(t) = \theta_1(\infty) * \lim_{t\to\infty} \alpha_k(t) = 0, 1 < k \le m, \tag{18}$$

так как последовательность λ_k положительная убывающая.

Случай 2. $k = 1$**.** Подставляя в (16) $k = 1$ получаем:

$$\frac{d\theta_1(t)}{dt} = \lambda_1\theta_1(t) - \left[\sum_{l=1}^{m} \lambda_l \theta_l^2(t)\right] \theta_1(t) = \lambda_1\theta_1(t) - \lambda_1\theta_1^3(t) - \left[\sum_{l=2}^{m} \lambda_l \theta_l^2(t)\right] \theta_1(t),$$

переходя к пределу $t \to \infty$ и воспользовавшись (18) окончательно получаем:

$$\frac{d\theta_1(t)}{dt} = \lambda_1\theta_1(t)[1 - \theta_1^2(t)] \text{ при } t \to \infty. \tag{19}$$

Функцией Ляпунова системы (19) будет $V(t) = [\theta_1^2(t) - 1]^2$. Это проверяется непосредственно по определению. Ее минимум достигается при

$$\theta_1(t) = \pm 1 \text{ при } t \to \infty. \tag{20}$$

Это и есть единственное асимптотически устойчивое решение (14).

В итоге пользуясь разложением (15) и предельными соотношениями (20), (18) получаем:

$$\mathbf{w}(t) \to \mathbf{q}_1 \text{ при } t \to \infty. \tag{21}$$

Таким образом мы удолетворили все условия теоремы 3, следовательно

$$\lim_{n\to\infty} \mathbf{w}(n) = \mathbf{q}_1 \text{ бесконечно часто с вероятностью 1.} \tag{22}$$

То есть

$$y(n) = \langle \mathbf{x}(n), \mathbf{w}(n) \rangle = \langle \mathbf{x}(n), \mathbf{q}_1 \rangle = \lambda_1, \text{ при } n \to \infty. \tag{23}$$

А мы знаем, что максимальное собственное число λ_1 матрицы корреляции $\mathbf{R} = E[\mathbf{X}\mathbf{X}^\mathbf{T}]$ есть первый главный компонент вектора $\mathbf{x}$.

9. Анализ главных компонент на основе фильтра Хебба

В предыдущих разделах мы научились извлекать первый главный компонент вектора $\mathbf{x} = \{x_i\}_{i=1}^m$. Этот алгоритм естественным образом обобщается для извлечения $l <= m$ главных компонент.

Пусть теперь у нас есть не 1 а l линейных нейронов, на вход каждому из которых подается вектор сигналов $\mathbf{x}$. Обозначим $\mathbf{y} = \{y_j\}_{j=1}^l$ вектор выходных сигналов, y_j — выход j-го нейрона. w_{ij} — вес i-го синапса j-го нейрона. $y_j(n) = \sum_{i=1}^m w_{ij}(n)x_i(n)$.

Можно показать (аналогично случаю одного нейрона), что если использовать правило обучения Хебба

$$\Delta w_{ij}(n) = \eta y_j(n) \left[x_i(n) - \sum_{k=1}^{j} w_{ki}(n) y_k(n) \right], \tag{24}$$

то

$$\lim_{n\to\infty} \mathbf{w}_j(n) = \mathbf{q}_j,$$

а следовательно, $y_j(n)$ будет стремиться к j-му главному компоненту $\mathbf{x}$.