

Точные оценки вероятности переобучения

Задачи по спецкурсу

«Теория надёжности обучения по прецедентам»

К. В. Воронцов (www.ccas.ru/voron)

10 мая 2010 г.

Точные комбинаторные оценки вероятности переобучения — это новое направление исследований в теории статистического обучения. В нём ещё очень много открытых проблем, многие из которых достаточно просты и вполне по силам студентам 3 курса. Чтобы понимать, о чём эти задачи, необходимо проработать статьи [1, 2], или диссертацию [3], или ходить на спецкурс.

О структуре данного документа. Обозначения и предположения, вводимые в начале каждого параграфа, распространяются на все задачи внутри данного параграфа. Перед каждым блоком задач даётся минимум необходимых пояснений, включая мотивации, которые привели к этим задачам.

О формальностях. Для сдачи спецкурса необходимо набрать определённое количество баллов, которое будет объявлено на лекциях. Стоимость каждой задачи в баллах указана в скобках после номера задачи.

Страница курса: вики-ресурс www.MachineLearning.ru/wiki, страница «Теория надёжности обучения по прецедентам (курс лекций, К.В.Воронцов)». Адрес этого документа: www.MachineLearning.ru/wiki/images/6/6b/Voron09problems.pdf.

Содержание

1	Классические оценки в слабой вероятностной аксиоматике	3
§1.1	Закон больших чисел	3
§1.2	Некоторые статистические критерии	3
§1.3	Оценки вероятности переобучения	3
2	Исследования свойства связности	4
§2.1	Цепочки, порождаемые линейными классификаторами	4
§2.2	Графы связности, порождаемые линейными классификаторами	4
3	Точные оценки вероятности переобучения	5
§3.1	Цепочки	5
§3.2	Семейства алгоритмов без расслоения	6
§3.3	Реальные семейства алгоритмов	6
4	Семейства, задаваемые перечислением векторов ошибок	6
§4.1	Оценки по ненаблюдаемым данным	7
§4.2	Оценки по наблюдаемым данным	7
5	Асимптотические оценки	8
6	Экспериментальные исследования (практикум)	8
§6.1	Эмпирическое исследование вероятности переобучения	8
§6.2	Визуализация графов связности	9

1 Классические оценки в слабой вероятностной аксиоматике

Задачи этой секции направлены на то, чтобы потренироваться в доказательстве оценок вероятности больших отклонений или вероятности переобучения, пользуясь только слабой вероятностной аксиоматикой.

Слабая вероятностная аксиоматика содержит единственную аксиому: дана конечная выборка $\mathbb{X} = \{x_1, \dots, x_L\}$, и все её разбиения $X \cup \bar{X} = \mathbb{X}$ на наблюдаемую (обучающую) выборку X длины ℓ и скрытую (контрольную) выборку \bar{X} длины k равновероятны.

§1.1 Закон больших чисел

Для любого алгоритма a , допускающего $m = n(a, \mathbb{X})$ ошибок на полной выборке, справедлива точная оценка вероятности большого отклонения частоты ошибок [4]:

$$\mathbb{P}[\delta(a, X) \geq \varepsilon] = \sum_{s=s_0}^{s_1(\varepsilon)} h_L^{\ell, m}(s) \equiv H_L^{\ell, m}(s_1(\varepsilon)). \quad (1.1)$$

Это выражение можно рассматривать как оценку скорости сходимости в законе больших чисел для слабой аксиоматики, если положить $\ell, k \rightarrow \infty$.

Задача 1.1. (5) Вывести из этой оценки классические неравенства Чернова, Бернштейна, Хёффдинга [5]. Рассмотреть различные асимптотики: (а) $\ell \rightarrow \infty$ при фиксированном k , (б) $k \rightarrow \infty$ при фиксированном ℓ , (в) $\ell, k \rightarrow \infty$ при $\ell/k = \text{const}$.

§1.2 Некоторые статистические критерии

Многие статистические критерии, предназначенные для проверки гипотез однородности, независимости, случайности, стационарности (критерии Смирнова, Вилкоксона, ранговые и перестановочные критерии и т. д.), выводятся именно комбинаторными методами. Их (пере)формулировка и (пере)доказательство в слабой аксиоматике практически очевидны.

Задача 1.2. (1) Критерий знаков.

Задача 1.3. (1) Критерий Вилкоксона-Манна-Уитни.

Задача 1.4. (3) Критерий серий Вальда-Вольфовица.

§1.3 Оценки вероятности переобучения

Следующие две задачи заключаются в том, чтобы аккуратно воспроизвести в слабой аксиоматике результаты Бакса и Силла, которые оценивали другой функционал (функционал равномерной сходимости). Позволяет ли слабая аксиоматика упростить доказательства? Улучшить оценку?

Задача 1.5. (3) Следуя [6], показать, что если множество векторов ошибок $\{(a)_{\mathbb{X}}: a \in \mathbb{A}\}$ кластеризуется по расстоянию Хэмминга на $S(r)$ кластерах радиуса r каждый, то

$$\mathbb{P}[\delta_{\mu}(X) \geq \varepsilon + \frac{r}{\ell}] \leq S(r) \cdot \max_{m=1, \dots, L} H_L^{\ell, m}(s_1(\varepsilon)).$$

Задача 1.6. (3) Следуя [7, 8] показать, что если семейство A связно, то

$$\mathbb{P}[\delta_\mu(X) \geq \varepsilon] \leq \frac{1}{\sqrt{\pi L}} |A| \max_{m=1, \dots, L} H_L^{\ell, m}(s_1(\varepsilon)).$$

2 Исследования свойства связности

Достаточно просто должны доказываться утверждения о том, что при непрерывном изменении вектора параметров вдоль непрерывной траектории почти всегда образуется цепочка векторов ошибок. Что значит «почти всегда», необходимо уточнить.

§2.1 Цепочки, порождаемые линейными классификаторами

Пусть \mathbb{X} — множество, состоящее из L точек в \mathbb{R}^n ; $\mathbb{Y} = \{-1, +1\}$ — множество меток классов; \mathbb{A} — семейство n -мерных линейных классификаторов:

$$\mathbb{A} = \{a: \mathbb{X} \rightarrow \mathbb{Y} \mid a(x; w) = \text{sign}(x_1 w_1 + \dots + x_n w_n), w \in \mathbb{R}^n\}, \quad (2.1)$$

где $w \in \mathbb{R}^n$ — вектор параметров. Пусть $I(a, x) = [a(x) \neq y(x)]$ — бинарная функция потерь, где $y: \mathbb{X} \rightarrow \mathbb{Y}$ — функция истинной классификации.

Задача 2.1. (1) Доказать, что множество векторов ошибок $\{a(x; w + t\delta): t \in \mathbb{R}\}$ при некоторых дополнительных ограничениях образует цепочку. Сформулировать эти ограничения. Какова максимальная возможная длина цепочки?

Говорят, что множество объектов \mathbb{X} линейно разделимо, если существует направляющий вектор разделяющей гиперплоскости $w^* \in \mathbb{R}^n$, при котором алгоритм $a(x; w^*)$ не допускает ошибок на \mathbb{X} .

Задача 2.2. (1) В каких случаях множество векторов ошибок $\{a(x; w^* + t\delta): t \geq 0\}$, где $\delta \in \mathbb{R}^n$ — фиксированный вектор, образует монотонную цепочку с $m = 0$?

Задача 2.3. (1) В каких случаях множество векторов ошибок $\{a(x; w^* + t\delta): t \in \mathbb{R}\}$ где $\delta \in \mathbb{R}^n$ — фиксированный вектор, образует унимодальную цепочку с $m = 0$?

§2.2 Графы связности, порождаемые линейными классификаторами

Рассматривается множество всех векторов ошибок, порождаемых линейными классификаторами (2.1) на заданной выборке \mathbb{X} в пространстве размерности $n = 2$.

В графе связности вершинами являются все векторы ошибок $\{\vec{a}: a \in \mathbb{A}\}$; рёбрами соединяются пары векторов с хэмминговым расстоянием 1.

Задача 2.4. (1) Привести пример выборки \mathbb{X} , для которой граф связности представляет собой двумерную решетку (у каждой вершины не более четырёх рёбер). Привести примеры выборок, для которых в графе связности имеются вершины с числом рёбер более четырёх.

Задача 2.5. (1) Доказать, что для любого L существует выборка длины L , для которой в графе связности имеется вершина с числом рёбер $L - 1$.

3 Точные оценки вероятности переобучения

Во всех задачах данного раздела задаётся множество векторов ошибок, и требуется выписать точные оценки P_a и Q_ε .

§3.1 Цепочки

Цепочки с фиксированным уровнем ошибок интересны тем, что это связанные семейства без расслоения. Сопоставляя их с монотонными цепочками, можно оценить влияние расслоения на вероятность переобучения.

Задача 3.1. (3) *Распрямлённая цепочка с уровнем ошибок m — это множество векторов ошибок a_0, a_1, \dots, a_D такое, что $n(a_d, \mathbb{X}) = m + (d \bmod 2)$ и хэммингово расстояние $\rho(a_{d'}, a_d) = d' - d$ для любых $d, d' = 0, \dots, D$.*

Задача 3.2. (5) *Скрученная цепочка с уровнем ошибок m — это множество векторов ошибок a_0, a_1, \dots, a_D такое, что $n(a_d, \mathbb{X}) = m + (d \bmod 2)$ и суммарное хэммингово расстояние $\sum_{d=1}^D \rho(a_0, a_d)$ минимально.*

Следующие две задачи аналогичны двум предыдущим, но все векторы ошибок находятся в одном слое, а не в двух соседних, и расстояние между соседними алгоритмами равно 2, а не 1.

Задача 3.3. (3) *Множество векторов ошибок образует m -диагональную матрицу векторов ошибок, то есть алгоритм a_d допускает ошибки на объектах x_{d+1}, \dots, x_{d+m} , $d = 0, \dots, D$.*

Задача 3.4. (5) *Множество векторов ошибок a_0, \dots, a_D такое, что $n(a_d, \mathbb{X}) = m$ и суммарное хэммингово расстояние $\sum_{d=1}^D \rho(a_0, a_d)$ минимально.*

В следующих задачах требуется выписать либо точную оценку, либо эффективный (не требующий суммирования по всем разбиениям выборки) алгоритм для её вычисления.

Задача 3.5. (10) *Произвольная цепочка алгоритмов без расслоения, $n(a_d, \mathbb{X}) = m + (d \bmod 2)$.*

Цепочки алгоритмов произвольного вида возникают на практике при выборе оптимального значения порога в решающих правилах алгоритмов классификации. Например, в линейных классификаторах вида $a(x; w) = \text{sign}(x_1 w_1 + \dots + x_n w_n - w_0)$ может потребоваться настроить порог w_0 при фиксированных w_1, \dots, w_n .

Задача 3.6. (3) *Привести пример цепочки (построить матрицу ошибок), для которой «очевидный» выбор порога по принципу минимума эмпирического риска может быть неоптимальным с точки зрения вероятности переобучения.*

Задача 3.7. (10) *Произвольная цепочка алгоритмов.*

Следующая задача — это попытка выяснить, монотонная цепочка какой длины способна «защитить» от переобучения, связанного с наличием полного слоя.

Задача 3.8. (5) Объединение монотонной цепочки a_m, \dots, a_M , $n(a_d, \mathbb{X}) = d$, и M -го слоя булева куба — множества всех алгоритмов $a: n(a, \mathbb{X}) = M$.

§3.2 Семейства алгоритмов без расслоения

Семейства алгоритмов без связности и без расслоения обладают наихудшими оценками вероятности переобучения.

В следующих задачах матрица ошибок не конкретизируется однозначно. Можно строить такую матрицу ошибок, при которой оценка получается наиболее просто.

Задача 3.9. (5) Множество алгоритмов a_1, \dots, a_D такое, что $n(a_d, \mathbb{X}) = m$, $\text{rho}(a_d, a_{d+1}) = r$, алгоритмы a_d, a_{d+i} , $i > 1$ различны максимально, насколько возможно.

В каких случаях достигается Вапниковская оценка вероятности переобучения?

Следующая задача связана с равновесными кодами, исправляющими ошибки.

Задача 3.10. (15) Множество алгоритмов a_1, \dots, a_D такое, что $n(a_d, \mathbb{X}) = m$, $\text{rho}(a_d, a_{d'}) \leq r$ для всех $d, d' = 1, \dots, D$.

В каких случаях достигается Вапниковская оценка вероятности переобучения?

§3.3 Реальные семейства алгоритмов

Задача 3.11. (10) Параметрическое семейство A линейных алгоритмов классификации над n вещественными признаками $f_1(x), \dots, f_n(x)$ с параметрами w_1, \dots, w_n :

$$a(x) = \text{sign} \left(\sum_{j=1}^n w_j f_j(x) \right).$$

Задача 3.12. (5) Предыдущая задача при $n = 2$.

Задача 3.13. (10) Параметрическое семейство A конъюнкций над n вещественными признаками $f_1(x), \dots, f_n(x)$: с параметрами w_1, \dots, w_n :

$$a(x) = \bigwedge_{j=1}^n [f_j(x) \leq w_j].$$

Задача 3.14. (5) Предыдущая задача при $n = 2$.

Задача 3.15. (20) Рассматривается семейство линейных алгоритмов классификации над n вещественными признаками. На норму вектора параметров накладывается дополнительное ограничение регуляризации $\|w\| \leq \tau$. Как при этом меняется вероятность переобучения?

4 Семейства, задаваемые перечислением векторов ошибок

В этой секции рассматриваются семейства A , заданные не слишком длинной совокупностью векторов ошибок, в общем случае произвольных. «Не слишком длинной» означает, что оценки должны вычисляться за приемлемое время при условии, что векторы ошибок в явном виде хранятся в памяти компьютера.

§4.1 Оценки по ненаблюдаемым данным

В следующих задачах векторы ошибок предполагаются известными полностью. На практике это невозможно, т.к. при фиксированном разбиении (X, \bar{X}) скрытая (контрольная) выборка неизвестна. Тем не менее, получение оценок Q_ε и в этих случаях представляет интерес. Во-первых, это позволит ускорить расчёты экспериментов, заменив эмпирические оценки по методу Монте-Карло точными оценками. Во-вторых, из этих оценок, возможно, будут проще получаться верхние оценки для некоторых семейств алгоритмов.

Простейшим примером такого явно заданного семейства является двухэлементное семейство, для которого точная оценка получена в [1, 2].

Задача 4.1. (3) Выписать точные оценки P_a и Q_ε для семейства из трёх произвольных алгоритмов $A = \{a_1, a_2, a_3\}$, считая, что заданы восемь параметров

$$m_{abc} = \#\{x \in \mathbb{X} : I(a_1, x) = a, I(a_2, x) = b, I(a_3, x) = c\}, \quad a, b, c \in \{0, 1\}.$$

Задача 4.2. (2) В задаче 4.1 рассмотреть частный случай — трёхэлементное семейство без расслоения: $m_{110} = m_{101} = m_{011} = m_{100} = m_{010} = m_{001} = m$. Построить зависимость Q_ε от хэммингова расстояния между векторами ошибок (которое равно $4m$) при нескольких (небольших) значениях m_{111} .

Задача 4.3. (2) В задаче 4.1 рассмотреть частный случай — трёхэлементное семейство с расслоением: $m_{011} = m_{001} = m$, $m_{110} = m_{101} = m_{100} = m_{010} = 0$. Построить зависимость Q_ε от хэммингова расстояния между векторами ошибок $\rho(a_1, a_2) = \rho(a_2, a_3) = m$ при нескольких (небольших) значениях m_{111} .

§4.2 Оценки по наблюдаемым данным

В этих задачах предполагается, что векторы ошибок известны не полностью, а только на наблюдаемой выборке X при заданном разбиении (X, \bar{X}) , причём их можно явным образом перебирать и использовать их данные в вычислениях. На практике возможность явного перебора реализуется не всегда, а только в переборных методах обучения. Примеры таких методов:

- выбор лучшей модели по отложенным данным (hold-out model selection);
- стохастический поиск (например, генетические алгоритмы);
- поиск информативных конъюнкций в логических классификаторах;
- отбор признаков в линейной регрессии с фиксированными коэффициентами.

Предполагается, что получение оценок вероятности переобучения позволит некоторым образом улучшить эти методы.

Начнём с простейшего случая — семейства из двух алгоритмов.

Задача 4.4. (4) Для двухэлементного семейства алгоритмов $A = \{a_1, a_2\}$ при некотором разбиении (X, \bar{X}) , выбранном случайно и равновероятно, в наблюдаемой подвыборке X оказалось:

- s_0 объектов, на которых ошиблись оба алгоритма;
- s_1 объектов, на которых ошибся только алгоритм a_1 ;
- s_2 объектов, на которых ошибся только алгоритм a_2 .

Оценить сверху вероятность переобучения Q_ε для метода минимизации эмпирического риска. Как Q_ε зависит от $s_1 + s_2$ — наблюдаемого расстояния между алгоритмами? Как Q_ε зависит от $|s_1 - s_2|$ — наблюдаемой величины расслоения алгоритмов?

Задача 4.5. (∞) Для d -элементного семейства алгоритмов $A = \{a_1, \dots, a_d\}$ при некотором разбиении (X, \bar{X}) , выбранном случайно и равновероятно, известны векторы ошибок $(a_1)_X, \dots, (a_d)_X$. Оценить сверху вероятность переобучения Q_ε для метода минимизации эмпирического риска. Как Q_ε зависит от расстояния между алгоритмами? Использование каких характеристик схожести алгоритмов позволяет записать более точную оценку Q_ε ?

5 Асимптотические оценки

Точные оценки вероятности переобучения имеют громоздкий вид и сложны для вычислений. Наверняка эти комбинаторные выражения можно упростить. В асимптотике $L \rightarrow \infty$ будем предполагать, что $\frac{\ell}{L} \rightarrow \text{const}$, $\frac{m}{L} \rightarrow \text{const}$.

Задача 5.1. (3) Найти асимптотическое выражение точной верхней оценки вероятности переобучения для монотонной цепочки.

Задача 5.2. (3) Найти асимптотическое выражение точной верхней оценки вероятности переобучения для унимодальной цепочки.

Задача 5.3. (3) Найти асимптотическое выражение точной верхней оценки вероятности переобучения для единичной окрестности лучшего алгоритма.

6 Экспериментальные исследования (практикум)

Реализация следующих программ и экспериментирование с ними поможет не только проверять правильность теоретических оценок, но и замечать новые интересные факты, выдвигать и проверять новые гипотезы.

Графики желательно генерировать в виде chd-файлов в формате ChartLib.

§6.1 Эмпирическое исследование вероятности переобучения

Задача 6.1. (5) Написать программу, позволяющую:

- генерировать модельные семейства алгоритмов $A = \{a_1, \dots, a_D\}$ в виде бинарной матрицы ошибок размера $L \times D$ (в матрице ошибок не должно быть одинаковых столбцов — векторов ошибок); легко заменять генераторы данных;
- генерировать векторы ошибок так, чтобы в A последовательно формировались новые слои, и все оценки вычислялись для каждого из подмножеств алгоритмов $A(m) = \{a_d : n(a_d, \mathbb{X}) \leq m\}$;
- вычислять точные верхние и нижние оценки вероятности переобучения, если соответствующие формулы известны;

- вычислять эмпирические оценки вероятности переобучения методом Монте-Карло, т. е. по случайному подмножеству из N разбиений (X, \bar{X}) ; требуется вычислять три оценки, соответствующие трём стратегиям выбора алгоритма в случаях неоднозначного минимума эмпирического риска:
 - верхняя оценка \bar{Q}_ε для пессимистичного $\mu_{\text{пес}}$ (худший из лучших);
 - нижняя оценка $\underline{Q}_\varepsilon$ для оптимистичного $\mu_{\text{опт}}$ (лучший из лучших);
 - средняя оценка \hat{Q}_ε для рандомизированного $\mu_{\text{ран}}$ (случайный из лучших).
- строить графики, откладывая по оси X номер слоя m , по оси Y :
 - эмпирические оценки $\bar{Q}_\varepsilon, \underline{Q}_\varepsilon, \hat{Q}_\varepsilon$ для подмножества $A(m)$;
 - точные значения $\bar{Q}_\varepsilon, \underline{Q}_\varepsilon, \hat{Q}_\varepsilon$ для подмножества $A(m)$ (если известны);
 - число алгоритмов в m -м слое;
 - доля разбиений, на которых $n(\mu_{\text{пес}}X, \mathbb{X}) = m$;
 - доля разбиений, на которых $n(\mu_{\text{опт}}X, \mathbb{X}) = m$.
- строить графики, в которых по оси Y откладываются точные (если известны) и эмпирические значения $\bar{Q}_\varepsilon, \underline{Q}_\varepsilon, \hat{Q}_\varepsilon$, по оси X :
 - число ошибок лучшего алгоритма m ;
 - длина обучения ℓ ;
 - длина контроля k .

Следующая серия задач направлена на экспериментальную проверку теоретических оценок и исследование зависимостей $\bar{Q}_\varepsilon, \underline{Q}_\varepsilon, \hat{Q}_\varepsilon$ от параметров модельного семейства алгоритмов. Число разбиений N должно быть достаточно большим ($10^3 \div 10^4$), чтобы совпадение теоретических и эмпирических оценок было очевидно.

Задача 6.2. (3) *Монотонная цепочка с параметрами m и D .*

Задача 6.3. (3) *Унимодальная цепочка с параметрами m и D .*

Задача 6.4. (3) *Единичная окрестность лучшего алгоритма с параметрами m и D .*

Задача 6.5. (8) *Интервал булева куба с параметрами m_1 и ранг r .*

Следующий эксперимент направлен на проверку гипотезы, что монотонная цепочка и цепочка случайных инверсий [1] ведут себя практически одинаково с точки зрения переобучения. Если это подтвердится, то можно будет ограничиться изучением монотонных цепочек, как достаточно точной модели реальных цепочек.

Задача 6.6. (3) *Сравнить, представив на одном графике, точные значения \bar{Q}_ε для монотонной цепочки, их эмпирические оценки и эмпирические оценки \bar{Q}_ε для цепочки случайных инверсий при одинаковых m . Увеличиваются ли различия между \bar{Q}_ε монотонной цепочки и цепочки случайных инверсий с ростом ℓ и m ?*

§6.2 Визуализация графов связности

Задача 6.7. (5) *Написать программу, позволяющую:*

- генерировать двумерные модельные задачи классификации на два класса;

- строить точечный график выборки;
- строить бинарную матрицу ошибок, порождаемую всевозможными двумерными линейными классификаторами на заданной выборке (в матрице ошибок не должно быть одинаковых столбцов — векторов ошибок);
- отображать граф связности в виде плоского точечного графика, со всеми рёбрами, желательно без самопересечений.

Задача 6.8. (3) Отобразить выборки и графы связности для Задач 2.4 и 2.5.

Задача 6.9. (10) Обобщить программу так, чтобы она допускала использование широкого класса методов обучения, а не только линейных классификаторов, в частности: ближайших соседей и других метрических классификаторов, двухслойных нейронных сетей, радиальных базисных функций, решающих деревьев, и т. п.

Список литературы

- [1] Vorontsov K. V. Splitting and similarity phenomena in the sets of classifiers and their effect on the probability of overfitting // *Pattern Recognition and Image Analysis*. — 2009. — Vol. 19, no. 3. — Pp. 412–420.
<http://www.MachineLearning.ru/wiki/images/0/0e/Voron09roai2008.pdf>.
- [2] Воронцов К. В. Точные оценки вероятности переобучения. — Вики-ресурс по машинному обучению, распознаванию образов и интеллектуальному анализу данных www.MachineLearning.ru. — 2009.
<http://www.MachineLearning.ru/wiki/images/1/18/Voron09exact.pdf>.
- [3] Воронцов К. В. Комбинаторная теория надёжности обучения по прецедентам. — Диссертация на соискание учёной степени д.ф.-м.н., М.: ВЦ РАН. — 2010.
<http://http://www.MachineLearning.ru/wiki/images/b/b6/Voron10doct.pdf>.
- [4] Воронцов К. В. Комбинаторная вероятность и точность оценок обобщающей способности (Combinatorial probability and the tightness of generalization bounds) // *Pattern Recognition and Image Analysis*. — 2008. — Vol. 18, no. 2. — Pp. 243–259.
<http://www.ccas.ru/frc/papers/voron08pria.pdf>.
- [5] Lugosi G. On concentration-of-measure inequalities. — Machine Learning Summer School, Australian National University, Canberra. — 2003.
<http://citeseer.ist.psu.edu/lugosi98concentrationmeasure.html>.
- [6] Bax E. T. Similar classifiers and VC error bounds: Tech. Rep. CalTech-CS-TR97-14: 1997.
<http://citeseer.ist.psu.edu/bax97similar.html>.
- [7] Sill J. Generalization bounds for connected function classes. — citeseer.ist.psu.edu/127284.html.
<http://citeseer.ist.psu.edu/127284.html>.

- [8] *Sill J.* Monotonicity and connectedness in learning systems: Ph.D. thesis / California Institute of Technology. — 1998.
<http://etd.caltech.edu/etd/available/etd-09222005-110351/>.