

БАЙЕСОВСКИЙ ВЫВОД ДЛЯ МОНЕТКИ

Сергей Николенко

uData School — Киев

22 мая 2018 г.

Random facts:

- 22 мая 853 г. Олаф Белый, сын короля Норвегии, принял в подчинение викингов и датчан в Ирландии и сделал Дублин её столицей
- 22 мая 1849 г. Авраам Линкольн получил патент на конструкцию плавучего дока (он единственный обладатель патента среди президентов США), 22 мая 1892 г. Вашингтон Шеффилд из Нью-Лондона запатентовал тюбик для зубной пасты, а 22 мая 1906 г. братья Райт получили патент на свой летательный аппарат
- 22 мая 1921 г. власти Чикаго приняли решение штрафовать на сумму от \$10 до \$100 женщин, появляющихся на улицах в коротких юбках и с обнажёнными руками
- 22 мая 1977 г. прошёл последний рейс «Восточного экспресса»

БАЙЕСОВСКИЙ ВЫВОД ДЛЯ МОНЕТКИ

- Мы остановились на том, что в статистике обычно ищут *гипотезу максимального правдоподобия* (maximum likelihood):

$$\theta_{ML} = \arg \max_{\theta} p(D | \theta).$$

- В байесовском подходе ищут *апостериорное распределение* (posterior)

$$p(\theta|D) \propto p(D|\theta)p(\theta)$$

и, возможно, *максимальную апостериорную гипотезу* (maximum a posteriori):

$$\theta_{MAP} = \arg \max_{\theta} p(\theta | D) = \arg \max_{\theta} p(D | \theta)p(\theta).$$

ПОСТАНОВКА ЗАДАЧИ

- Простая задача вывода: дана нечестная монетка, она подброшена N раз, имеется последовательность результатов падения монетки. Надо определить её «нечестность» и предсказать, чем она выпадет в следующий раз.

ПОСТАНОВКА ЗАДАЧИ

- Если у нас есть вероятность p_h того, что монетка выпадет решкой (вероятность орла $p_t = 1 - p_h$), то вероятность того, что выпадет последовательность s , которая содержит n_h решек и n_t орлов, равна

$$p(s|p_h) = p_h^{n_h} (1 - p_h)^{n_t}.$$

- Сделаем предположение: будем считать, что монетка выпадает равномерно, т.е. у нас нет априорного знания p_h .
- Теперь нужно использовать теорему Байеса и вычислить скрытые параметры.

ПРИМЕР ПРИМЕНЕНИЯ ТЕОРЕМЫ БАЙЕСА

- Правдоподобие: $p(p_h | s) = \frac{p(s|p_h)p(p_h)}{p(s)}$.
- Здесь $p(p_h)$ следует понимать как непрерывную случайную величину, сосредоточенную на интервале $[0, 1]$, коей она и является. Наше предположение о равномерном распределении в данном случае значит, что априорная вероятность $p(p_h) = 1, p_h \in [0, 1]$ (т.е. априори мы не знаем, насколько нечестна монетка, и предполагаем это равновероятным). А $p(s|p_h)$ мы уже знаем.
- Итого получается:

$$p(p_h | s) = \frac{p_h^{n_h} (1 - p_h)^{n_t}}{p(s)}$$

- Итого получается:

$$p(p_h | s) = \frac{p_h^{n_h} (1 - p_h)^{n_t}}{p(s)}.$$

- $p(s)$ можно подсчитать как

$$\begin{aligned} p(s) &= \int_0^1 p_h^{n_h} (1 - p_h)^{n_t} dp_h = \\ &= \frac{\Gamma(n_h + 1)\Gamma(n_t + 1)}{\Gamma(n_h + n_t + 2)} = \frac{n_h! n_t!}{(n_h + n_t + 1)!}, \end{aligned}$$

но найти $\arg \max_{p_h} p(p_h | s) = \frac{n_h}{n_h + n_t}$ можно и без этого.

ПРИМЕР ПРИМЕНЕНИЯ ТЕОРЕМЫ БАЙЕСА

- Итого получается:

$$p(p_h|s) = \frac{p_h^{n_h} (1 - p_h)^{n_t}}{p(s)}.$$

- Но это ещё не всё. Чтобы предсказать следующий исход, надо найти $p(\text{heads}|s)$:

$$\begin{aligned} p(\text{heads}|s) &= \int_0^1 p(\text{heads}|p_h) p(p_h|s) dp_h = \\ &= \int_0^1 \frac{p_h^{n_h+1} (1 - p_h)^{n_t}}{p(s)} dp_h = \\ &= \frac{(n_h + 1)! n_t!}{(n_h + n_t + 2)!} \cdot \frac{(n_h + n_t + 1)!}{n_h! n_t!} = \frac{n_h + 1}{n_h + n_t + 2}. \end{aligned}$$

- Получили правило Лапласа.

- Итого получается:

$$p(p_h|s) = \frac{p_h^{n_h}(1-p_h)^{n_t}}{p(s)}.$$

- Это была иллюстрация двух основных задач байесовского вывода:
 - найти апостериорное распределение на гипотезах/параметрах:

$$p(\theta | D) \propto p(D|\theta)p(\theta)$$

(и/или найти гипотезу максимального правдоподобия $\arg \max_{\theta} p(\theta | D)$);

- найти апостериорное распределение исходов дальнейших экспериментов:

$$p(x | D) \propto \int_{\theta \in \Theta} p(x | \theta)p(D|\theta)p(\theta)d\theta.$$

- Напоминаю, что основная наша задача – как обучить параметры распределения и/или предсказать следующие его точки по имеющимся данным.
- В байесовском выводе участвуют:
 - $p(x | \theta)$ – правдоподобие данных;
 - $p(\theta)$ – априорное распределение;
 - $p(x) = \int_{\Theta} p(x | \theta)p(\theta)d\theta$ – маргинальное правдоподобие;
 - $p(\theta | x) = \frac{p(x|\theta)p(\theta)}{p(x)}$ – апостериорное распределение;
 - $p(x' | x) = \int_{\Theta} p(x' | \theta)p(\theta | x)d\theta$ – предсказание нового x' .
- Задача обычно в том, чтобы найти $p(\theta | x)$ и/или $p(x' | x)$.

- Когда мы проводим байесовский вывод, у нас, кроме правдоподобия, должно быть ещё *априорное распределение* (prior distribution) по всем возможным значениям параметров.
- Мы раньше к ним специально не присматривались, но они очень важны.
- Задача байесовского вывода – как подсчитать $p(\theta | x)$ и/или $p(x' | x)$.
- Но чтобы это сделать, сначала надо выбрать $p(\theta)$. Как выбирать априорные распределения?

- Разумная цель: давайте будем выбирать распределения так, чтобы они оставались такими же и *a posteriori*.
- До начала вывода есть априорное распределение $p(\theta)$.
- После него есть какое-то новое апостериорное распределение $p(\theta | x)$.
- Я хочу, чтобы $p(\theta | x)$ тоже имело тот же вид, что и $p(\theta)$, просто с другими параметрами.

- Не слишком формальное определение: семейство распределений $p(\theta | \alpha)$ называется семейством *сопряжённых априорных распределений* для семейства правдоподобий $p(x | \theta)$, если после умножения на правдоподобие апостериорное распределение $p(\theta | x, \alpha)$ остаётся в том же семействе: $p(\theta | x, \alpha) = p(\theta | \alpha')$.
- α называются *гиперпараметрами* (hyperparameters), это «параметры распределения параметров».
- Тривиальный пример: семейство всех распределений будет сопряжённым чему угодно, но это не очень интересно.

- Разумеется, вид хорошего априорного распределения зависит от вида распределения собственно данных, $p(x | \theta)$.
- Сопряжённые априорные распределения подсчитаны для многих распределений, мы приведём несколько примеров.

- Каким будет сопряжённое априорное распределение для бросания нечестной монетки (испытаний Бернулли)?
- Ответ: это будет бета-распределение; плотность распределения нечестности монетки θ

$$p(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}.$$

- Плотность распределения нечестности монетки θ

$$p(\theta | \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}.$$

- Тогда, если мы посэмплируем монетку, получив s орлов и f решек, получится

$$p(s, f | \theta) = \binom{s+f}{s} \theta^s (1-\theta)^f, \text{ и}$$

$$\begin{aligned} p(\theta | s, f) &= \frac{\binom{s+f}{s} \theta^{s+\alpha-1} (1-\theta)^{f+\beta-1} / B(\alpha, \beta)}{\int_0^1 \binom{s+f}{s} x^{s+\alpha-1} (1-x)^{f+\beta-1} / B(\alpha, \beta) dx} = \\ &= \frac{\theta^{s+\alpha-1} (1-\theta)^{f+\beta-1}}{B(s+\alpha, f+\beta)}. \end{aligned}$$

- Итого получается, что сопряжённое априорное распределение для параметра нечестной монетки θ – это

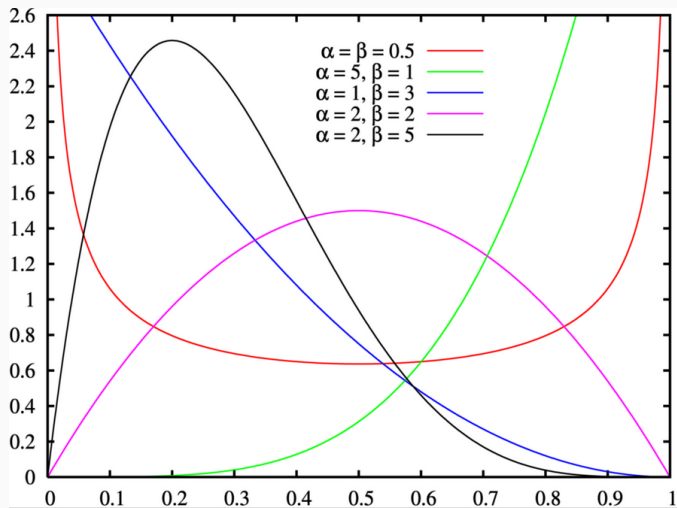
$$p(\theta \mid \alpha, \beta) \propto \theta^{\alpha-1}(1 - \theta)^{\beta-1}.$$

- После получения новых данных с s орлами и f решками гиперпараметры меняются на

$$p(\theta \mid s + \alpha, f + \beta) \propto \theta^{s+\alpha-1}(1 - \theta)^{f+\beta-1}.$$

- На этом этапе можно забыть про сложные формулы и выводы, получилось очень простое правило обучения (под обучением теперь понимается изменение гиперпараметров).

БЕТА--РАСПРЕДЕЛЕНИЕ



- Простое обобщение: рассмотрим мультиномиальное распределение с n испытаниями, k категориями и по x_i экспериментов дали категорию i .
- Параметры θ_i показывают вероятность попасть в категорию i :

$$p(x | \theta) = \binom{n}{x_1, \dots, x_n} \theta_1^{x_1} \theta_2^{x_2} \dots \theta_k^{x_k}.$$

- Сопряжённым априорным распределением будет распределение Дирихле:

$$p(\theta | \alpha) \propto \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} \dots \theta_k^{\alpha_k - 1}.$$

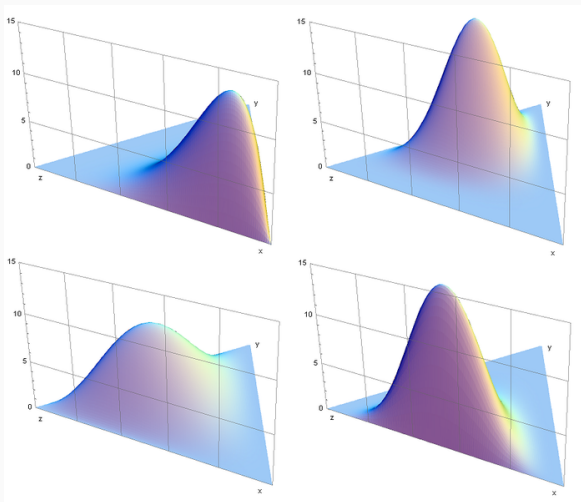
- Сопряжённым априорным распределением будет распределение Дирихле:

$$p(\theta | \alpha) \propto \theta_1^{\alpha_1-1} \theta_2^{\alpha_2-1} \dots \theta_k^{\alpha_k-1}.$$

Упражнение. Докажите, что при получении данных x_1, \dots, x_k гиперпараметры изменятся на

$$p(\theta | x, \alpha) = p(\theta | x + \alpha) \propto \theta_1^{x_1+\alpha_1-1} \theta_2^{x_2+\alpha_2-1} \dots \theta_k^{x_k+\alpha_k-1}.$$

РАСПРЕДЕЛЕНИЕ ДИРИХЛЕ



ЛИНЕЙНАЯ РЕГРЕССИЯ

- Линейная регрессия: рассмотрим линейную функцию

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^p x_j w_j = \mathbf{x}^\top \mathbf{w}, \quad \mathbf{x} = (1, x_1, \dots, x_p).$$

- Таким образом, по вектору входов $\mathbf{x}^\top = (x_1, \dots, x_p)$ мы будем предсказывать выход y как

$$\hat{y}(\mathbf{x}) = \hat{w}_0 + \sum_{j=1}^p x_j \hat{w}_j = \mathbf{x}^\top \hat{\mathbf{w}}.$$

- Как найти оптимальные параметры $\hat{\mathbf{w}}$ по тренировочным данным вида $(\mathbf{x}_i, y_i)_{i=1}^N$?
- Метод наименьших квадратов: будем минимизировать

$$\text{RSS}(\mathbf{w}) = \sum_{i=1}^N (y_i - \mathbf{x}_i^\top \mathbf{w})^2.$$

- Как минимизировать?

- Можно на самом деле решить задачу точно – записать как

$$\text{RSS}(\mathbf{w}) = (\mathbf{y} - \mathbf{X}\mathbf{w})^\top (\mathbf{y} - \mathbf{X}\mathbf{w}),$$

где \mathbf{X} – матрица $N \times p$, продифференцировать по \mathbf{w} , получится

$$\hat{\mathbf{w}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y},$$

если матрица $\mathbf{X}^\top \mathbf{X}$ невырожденная.

- Замечание: $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ называется *псевдообратной матрицей Мура–Пенроуза* (Moore–Penrose pseudo-inverse) матрицы \mathbf{X} ; это обобщение понятия обратной матрицы на неквадратные матрицы.
- Много ли нужно точек, чтобы обучить такую модель?

- Теперь давайте поговорим о линейной регрессии по-байесовски.
- Основное наше предположение – в том, что шум (ошибка в данных) распределён нормально, т.е. переменная t , которую мы наблюдаем, получается как

$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2).$$

Иными словами,

$$p(t \mid \mathbf{x}, \mathbf{w}, \sigma^2) = \mathcal{N}(t \mid y(\mathbf{x}, \mathbf{w}), \sigma^2).$$

- Здесь пока y – любая функция.

- Чтобы не повторять совсем уж то же самое, мы рассмотрим не в точности линейную регрессию, а её естественное обобщение – линейную модель с базисными функциями:

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x})$$

(M параметров, $M - 1$ базисная функция, $\phi_0(\mathbf{x}) = 1$).

- Базисные функции ϕ_i – это, например:
 - результат feature extraction;
 - расширение линейной модели на нелинейные зависимости (например, $\phi_j(x) = x^j$);
 - локальные функции, которые существенно не равны нулю только в небольшой области (например, гауссовские базисные функции $\phi_j(\mathbf{x}) = e^{-\frac{(x-\mu_j)^2}{2s^2}}$);
 - ...

- Рассмотрим набор данных $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ со значениями $\mathbf{t} = \{t_1, \dots, t_N\}$.
- Будем предполагать, что данные взяты независимо по одному и тому же распределению:

$$p(\mathbf{t} | \mathbf{X}, \mathbf{w}, \sigma^2) = \prod_{n=1}^N \mathcal{N}(t_n | \mathbf{w}^\top \phi(\mathbf{x}_n), \sigma^2).$$

- Прологарифмируем (опустим \mathbf{X} , т.к. по нему всегда условная вероятность будет):

$$\ln p(\mathbf{t} | \mathbf{w}, \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^N (t_n - \mathbf{w}^\top \phi(\mathbf{x}_n))^2.$$

- Прологарифмируем (опустим \mathbf{X} , т.к. по нему всегда условная вероятность будет):

$$\ln p(\mathbf{t} | \mathbf{w}, \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^N (t_n - \mathbf{w}^\top \phi(\mathbf{x}_n))^2.$$

- И вот мы получили, что для максимизации правдоподобия по \mathbf{w} нам нужно как раз минимизировать среднеквадратичную ошибку!

$$\nabla_{\mathbf{w}} \ln p(\mathbf{t} | \mathbf{w}, \sigma^2) = \frac{1}{\sigma^2} \sum_{n=1}^N (t_n - \mathbf{w}^\top \phi(\mathbf{x}_n)) \phi(\mathbf{x}_n).$$

- Решая систему уравнений $\nabla \ln p(\mathbf{t} | \mathbf{w}, \sigma^2) = 0$, получаем то же самое, что и раньше:

$$\mathbf{w}_{ML} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{t}.$$

- Здесь $\Phi = (\phi_j(\mathbf{x}_i))_{i,j}$.

- Теперь можно и относительно σ^2 максимизировать правдоподобие; получим

$$\sigma_{ML}^2 = \frac{1}{N} \sum_{n=1}^N (t_n - \mathbf{w}_{ML}^\top \phi(\mathbf{x}_n))^2,$$

т.е. как раз выборочная дисперсия имеющихся данных вокруг предсказанного значения.

ОВЕРФИТТИНГ В ЛИНЕЙНОЙ РЕГРЕССИИ

- Мы говорили о регрессии с базисными функциями:

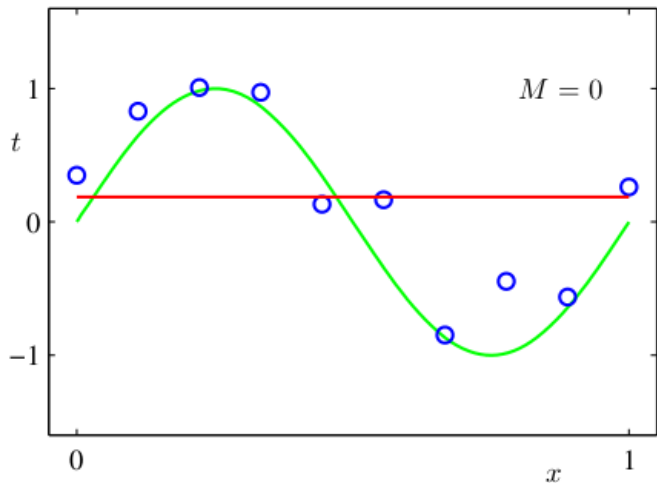
$$f(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^M w_j \phi_j(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x}).$$

- Давайте для примера рассмотрим такую регрессию для $\phi_j(x) = x^j$, т.е.

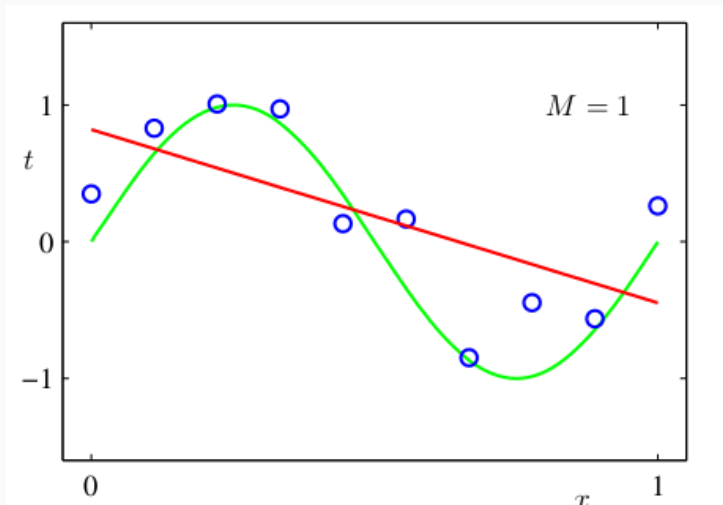
$$f(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \dots + w_M x^M.$$

- И будем, как раньше, минимизировать квадратичную ошибку.
- Пример с кодом.

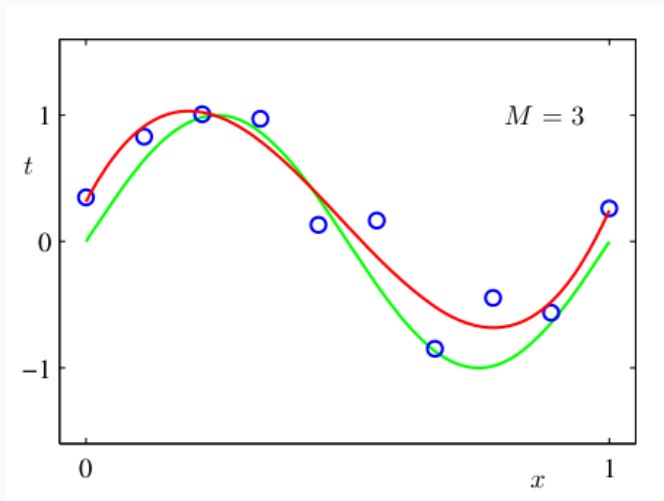
Полиномиальная аппроксимация



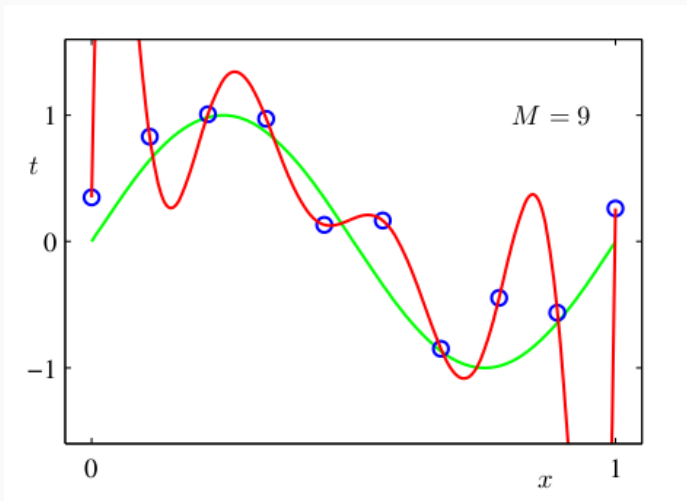
Полиномиальная аппроксимация



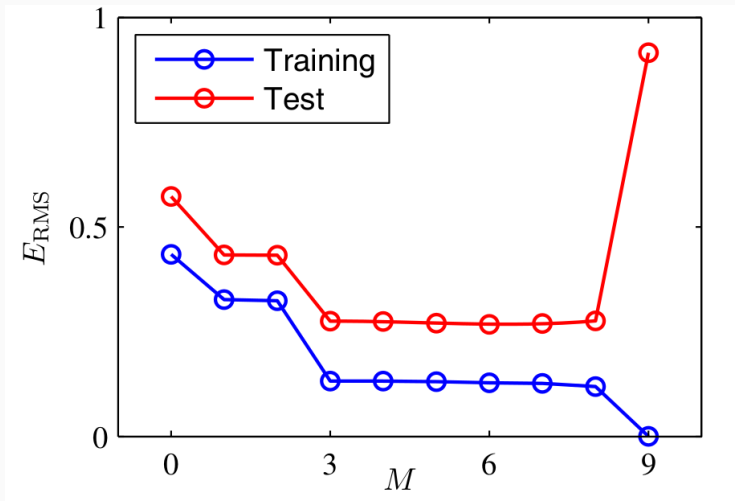
Полиномиальная аппроксимация



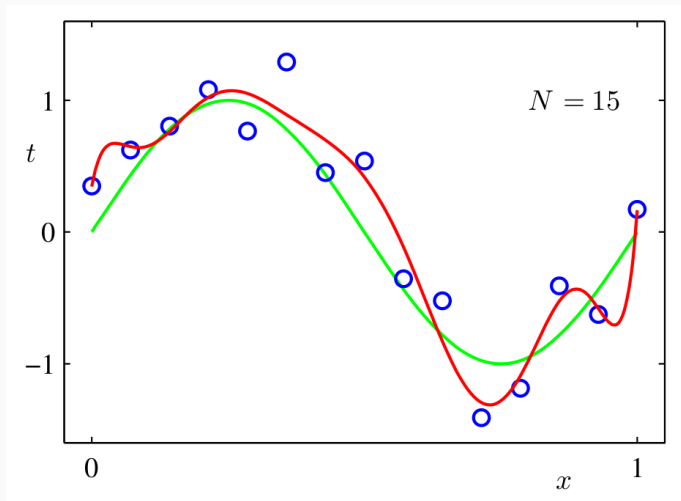
Полиномиальная аппроксимация



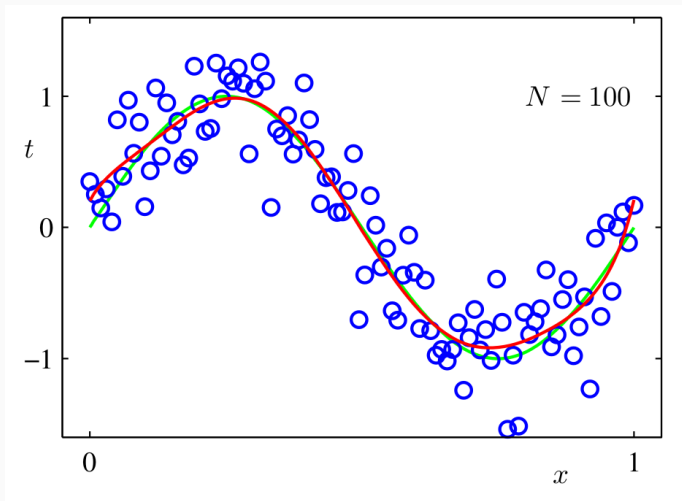
ЗНАЧЕНИЯ RMS



МОЖНО СОБРАТЬ БОЛЬШЕ ДАННЫХ...



МОЖНО СОБРАТЬ БОЛЬШЕ ДАННЫХ...



ЗНАЧЕНИЯ КОЭФФИЦИЕНТОВ

	$M = 0$	$M = 1$	$M = 6$	$M = 9$
w_0^*	0.19	0.82	0.31	0.35
w_1^*		-1.27	7.99	232.37
w_2^*			-25.43	-5321.83
w_3^*			17.37	48568.31
w_4^*				-231639.30
w_5^*				640042.26
w_6^*				-1061800.52
w_7^*				1042400.18
w_8^*				-557682.99
w_9^*				125201.43

- Итак, мы увидели, что даже в линейной регрессии может наступить оверфиттинг.
- Что же делать?..

РЕГУЛЯРИЗАЦИЯ В ЛИНЕЙНОЙ РЕГРЕССИИ

- Итак, получается, что у нас сильно растут коэффициенты.
- Давайте попробуем с этим бороться. Бороться будем прямолинейно и простодушно: возьмём и добавим размер коэффициентов в функцию ошибки.

- Было (для тестовых примеров $\{(x_i, y_i)\}_{i=1}^N$):

$$\text{RSS}(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (f(x_i, \mathbf{w}) - y_i)^2.$$

- Стало:

$$\text{RSS}(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (f(x_i, \mathbf{w}) - y_i)^2 + \frac{\alpha}{2} \|\mathbf{w}\|^2,$$

где α – коэффициент регуляризации (его надо будет как-нибудь выбрать).

- Как оптимизировать эту функцию ошибки?

- Да точно так же – запишем как

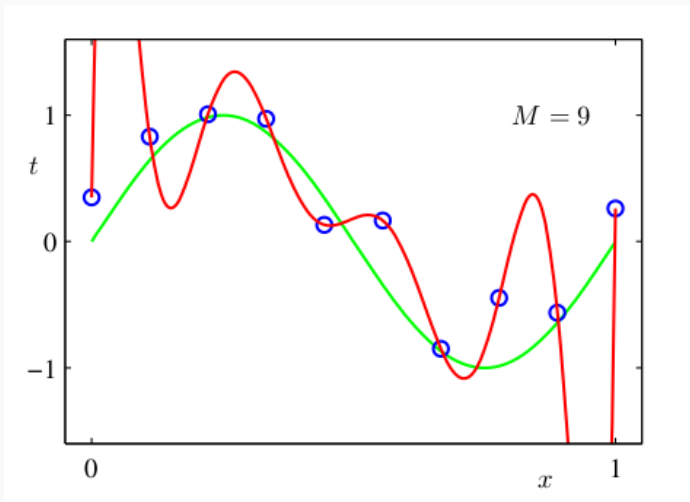
$$\text{RSS}(\mathbf{w}) = \frac{1}{2} (\mathbf{y} - \mathbf{X}\mathbf{w})^\top (\mathbf{y} - \mathbf{X}\mathbf{w}) + \frac{\alpha}{2} \mathbf{w}^\top \mathbf{w}$$

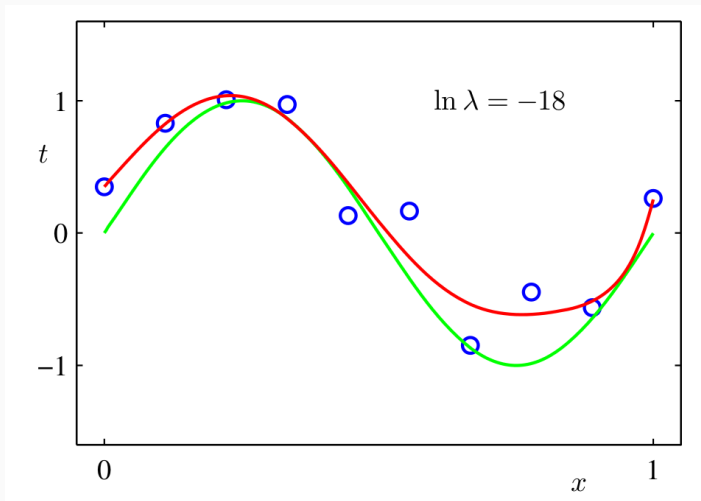
и возьмём производную; получится

$$\mathbf{w}^* = (\mathbf{X}^\top \mathbf{X} + \alpha \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}.$$

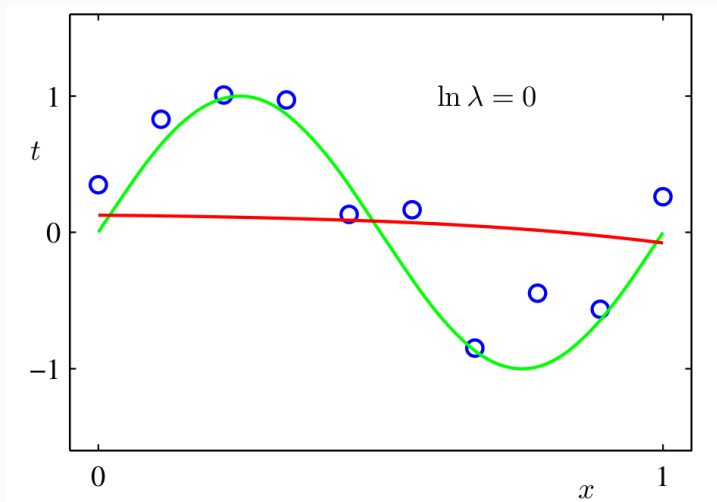
- Это *гребневая регрессия* (ridge regression); кстати, добавление $\alpha \mathbf{I}$ к матрице неполного ранга делает её обратимой; это и есть *регуляризация*, и это и было исходной мотивацией для гребневой регрессии.

ГРЕБНЕВАЯ РЕГРЕССИЯ: $\ln \alpha = -\infty$



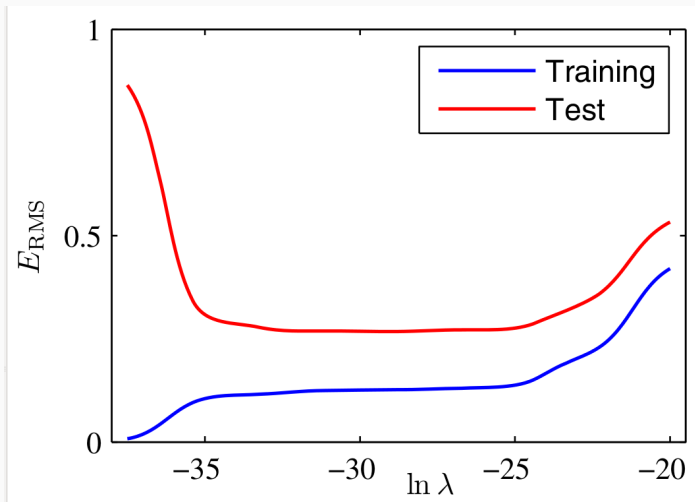


ГРЕБНЕВАЯ РЕГРЕССИЯ: $\ln \alpha = 0$



ГРЕБНЕВАЯ РЕГРЕССИЯ: КОЭФФИЦИЕНТЫ

	$\ln \lambda = -\infty$	$\ln \lambda = -18$	$\ln \lambda = 0$
w_0^*	0.35	0.35	0.13
w_1^*	232.37	4.74	-0.05
w_2^*	-5321.83	-0.77	-0.06
w_3^*	48568.31	-31.97	-0.05
w_4^*	-231639.30	-3.89	-0.03
w_5^*	640042.26	55.28	-0.02
w_6^*	-1061800.52	41.32	-0.01
w_7^*	1042400.18	-45.95	-0.00
w_8^*	-557682.99	-91.53	0.00
w_9^*	125201.43	72.68	0.01



- Почему именно так? Почему именно $\frac{\alpha}{2} \|\mathbf{w}\|^2$?
- Мы сейчас ответим на этот вопрос, но, вообще говоря, это не обязательно.
- Лассо-регрессия (lasso regression) регуляризует L_1 -нормой, а не L_2 :

$$\text{RSS}(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (f(x_i, \mathbf{w}) - y_i)^2 + \alpha \sum_{j=0}^M |w_j|.$$

- Есть и другие типы; об этом будем говорить позже.

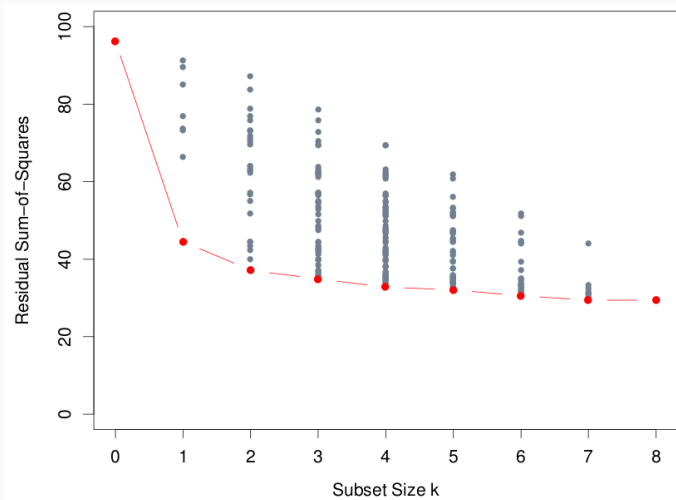
РАЗНЫЕ РЕГУЛЯРИЗАТОРЫ

- Мы знаем, что наименьшие квадраты не всегда хорошо работают. Две причины:
 1. плохая предсказательная сила – часто лучше регуляризовать, пожертвовав bias 'ом в пользу variance ;
 2. сложности в интерпретации – хотелось бы понимать, что происходит, если переменных с ненулевыми коэффициентами слишком много, не получится.
- Мораль: хотелось бы сделать так, чтобы было поменьше ненулевых компонент в векторе \mathbf{w} .

- Может быть, давайте так и сделаем? Будем искать самые лучшие компоненты и делать их ненулевыми.
- Это называется subset selection.
- Можно просто делать best subset selection: выбирать подмножество из k входных переменных, которые дают самые лучшие результаты.

- Это долго, даже если делать с умом, потому что subsets много.
- Forward-stepwise selection: начинаем со свободного члена, потом добавляет на каждом шаге предиктор, который максимально уменьшает ошибку.
- Т.е. подмножества тут получаются вложенные.
- Backward-stepwise selection: начинаем с полной регрессии и на каждом шаге убираем предиктор, который оказывает меньше всего влияния на ошибку.

SUBSET SELECTION



- Теперь давайте рассмотрим лассо-регрессию:

$$L(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (f(x_i, \mathbf{w}) - y_i)^2 + \lambda \sum_{j=0}^p |w_j|.$$

- Главное отличие – теперь форма ограничений (т.е. форма априорного распределения) такова, что весьма вероятно получить строго нулевые w_j .
- Кстати, что значит «форма ограничений»?

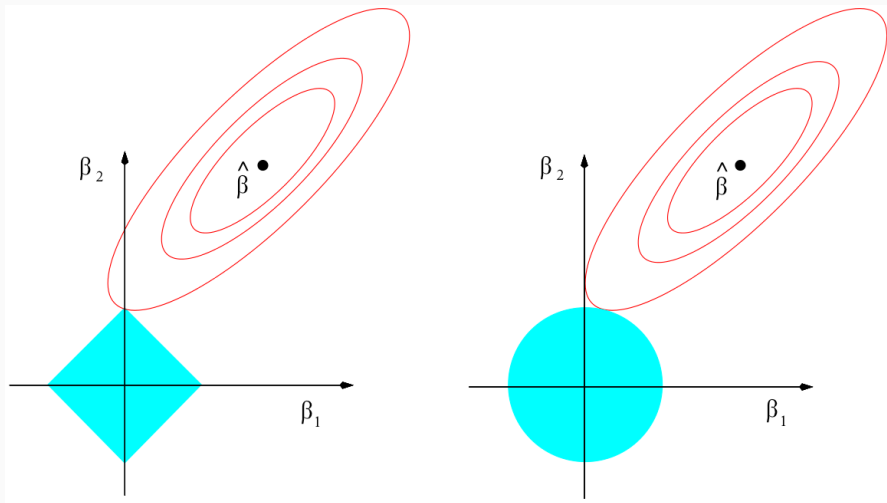
- Мы можем переписать регрессию с регуляризатором по-другому:

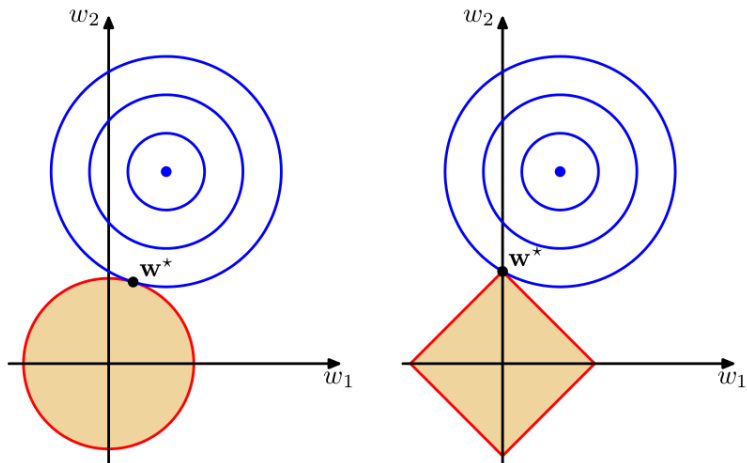
$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \left\{ \frac{1}{2} \sum_{i=1}^N (f(x_i, \mathbf{w}) - y_i)^2 + \lambda \sum_{j=0}^p |w_j| \right\},$$

эквивалентно

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \left\{ \frac{1}{2} \sum_{i=1}^N (f(x_i, \mathbf{w}) - y_i)^2 \right\} \text{ при } \sum_{j=0}^p |w_j| \leq t.$$

Упражнение. Докажите это.



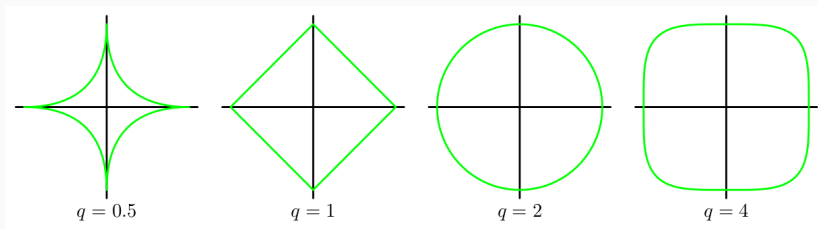
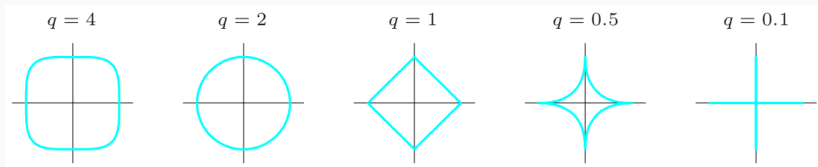


- Можно рассмотреть обобщение гребневой и лассо-регрессии:

$$L(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (f(x_i, \mathbf{w}) - y_i)^2 + \lambda \sum_{j=0}^p (|w_j|)^q.$$

Упражнение. Какому априорному распределению на параметры \mathbf{w} соответствует эта задача?

РАЗНЫЕ q



Спасибо за внимание!