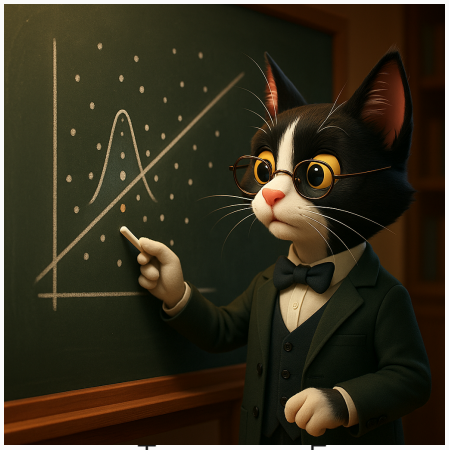
ЛИНЕЙНАЯ РЕГРЕССИЯ ПО-БАЙЕСОВСКИ



Сергей Николенко

СПбГУ — Санкт-Петербург

25 сентября 2025 г.

Random facts:

- 25 сентября 1066 г. Гаральд Гардерада вместе с Тостигом Годвинсоном потерпел сокрушительное поражение при Стэмфорд-Бридж, и попытка норвежского завоевания Англии провалилась
- 25 сентября 1303 г. произошло одно из самых смертоносных землетрясений в истории Земли: в Хундуне, в провинции Шаньси; погибли более 200 тысяч человек, землетрясение сровняло с землёй горы и холмы
- 25 сентября 1493 г. Христофор Колумб отправился в своё второе путешествие к Америке, а 25 сентября 1513 г. Васко де Бальбоа со своим отрядом пересек Панамский перешеек и стал первым европейцем, достигшим Тихого океана
- 25 сентября 1818 г. английский врач Джеймс Бландел впервые провёл операцию по переливанию крови от человека к человеку
- 25 сентября 1962 г. Фидель Кастро заявил, что СССР намерен создать на Кубе базу для своего флота; рыболовного, разумеется
- 25 сентября 1968 г. в первый и пока единственный раз британский хит-парад возглавила русская песня — романс «Дорогой длинною»; правда, называлась она «Those Were the Days» и исполнялась Мэри Хопкин со словами Джина Раскина

Регуляризация по-байесовски

Байесовская регуляризация

- А теперь давайте посмотрим на регрессию с совсем байесовской стороны.
- Напомним основу байесовского подхода:
  1. найти апостериорное распределение на гипотезах/параметрах:

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

     (и/или найти максимальную апостериорную гипотезу $\arg\max_\theta p(\theta \mid D)$);
  2. найти апостериорное распределение исходов дальнейших экспериментов:

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

Байесовская регуляризация

- В нашем рассмотрении пока не было никаких априорных распределений.
- Давайте какое-нибудь введём; например, нормальное (почему так – позже):

$$p(\mathbf{w}) = N(\mathbf{w} \mid \mu_0, \Sigma_0).$$

- Рассмотрим набор данных $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ со значениями $\mathbf{t} = \{t_1, \dots, t_N\}$. В этой модели мы предполагаем, что данные независимы и одинаково распределены:

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

Байесовская регуляризация

- Тогда наша задача – посчитать

$$p(\mathbf{w} \mid \mathbf{t}) \propto p(\mathbf{t} \mid \mathbf{X}, \mathbf{w}, \sigma^2) p(\mathbf{w})$$
$$= N(\mathbf{w} \mid \mu_0, \Sigma_0) \prod_{n=1}^{N} N\left(t_n \mid \mathbf{w}^\top \phi(\mathbf{x}_n), \sigma^2\right).$$

- Давайте подсчитаем.

Байесовская регуляризация

- Получится

$$p(\mathbf{w} \mid \mathbf{t}) = N(\mathbf{w} \mid \mu_N, \Sigma_N),$$
$$\mu_N = \Sigma_N \left(\Sigma_0^{-1} \mu_0 + \frac{1}{\sigma^2} \Phi^\top \mathbf{t} \right),$$
$$\Sigma_N = \left(\Sigma_0^{-1} + \frac{1}{\sigma^2} \Phi^\top \Phi \right)^{-1}.$$

- Теперь давайте подсчитаем логарифм правдоподобия.

Байесовская регуляризация

- Если мы возьмём априорное распределение около нуля:

$$p(\mathbf{w}) = N(\mathbf{w} \mid 0, \frac{1}{\alpha}\mathbf{I}),$$

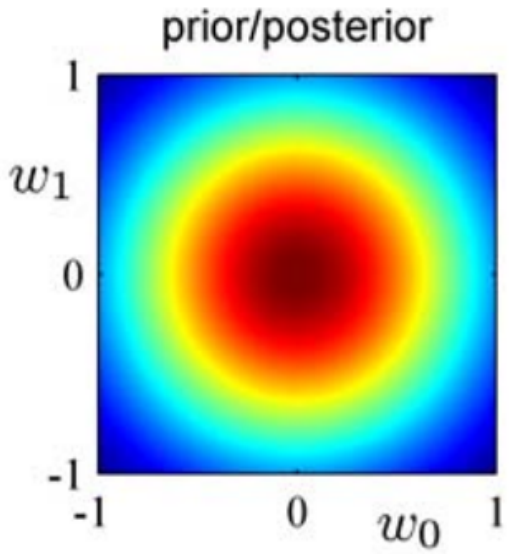
то логарифм правдоподобия получится

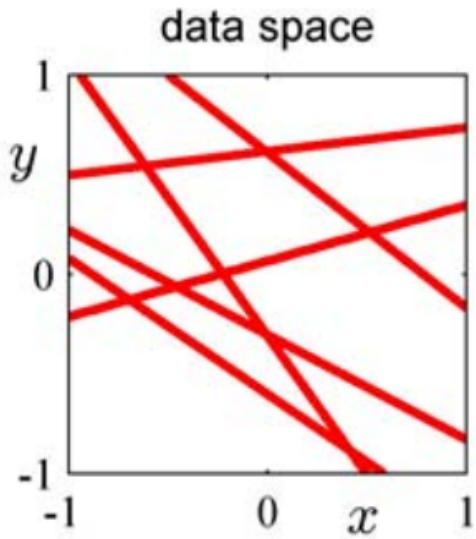
$$\ln p(\mathbf{w} \mid \mathbf{t}) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(t_n - \mathbf{w}^\top\phi(\mathbf{x}_n)\right)^2 - \frac{\alpha}{2}\mathbf{w}^\top\mathbf{w} + \text{const},$$

то есть в точности гребневая регрессия.

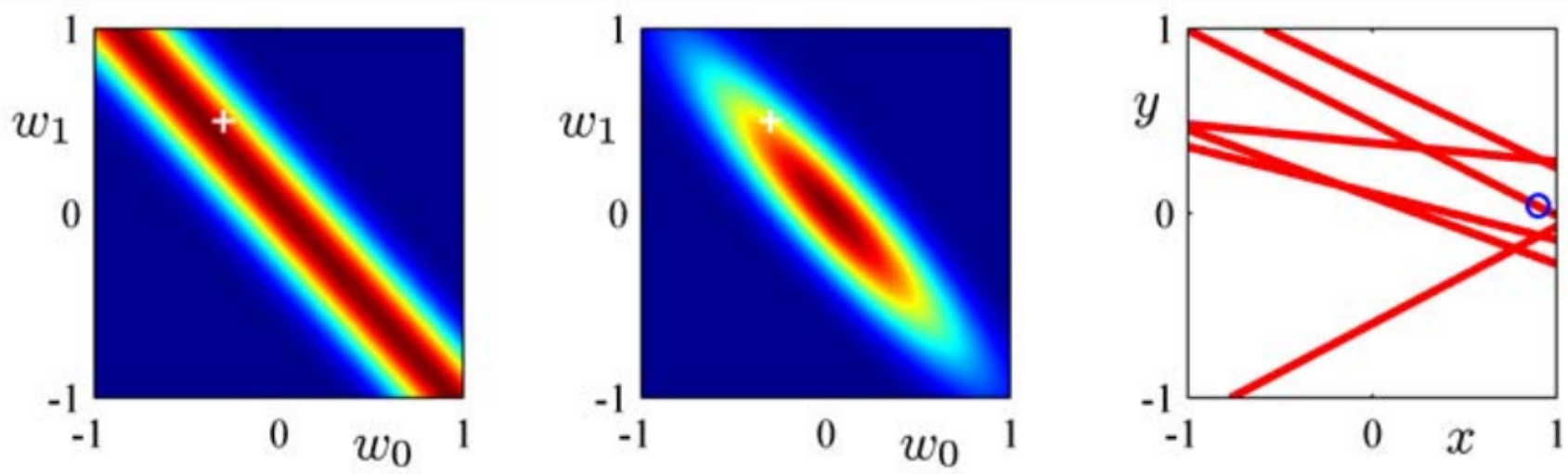
Пример

likelihood

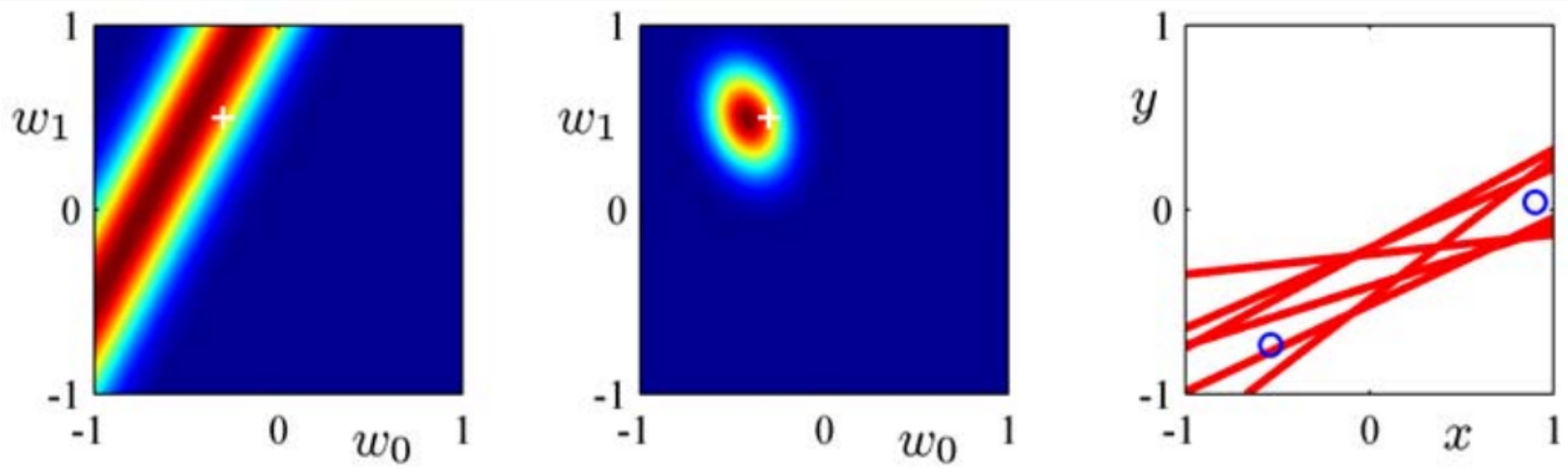




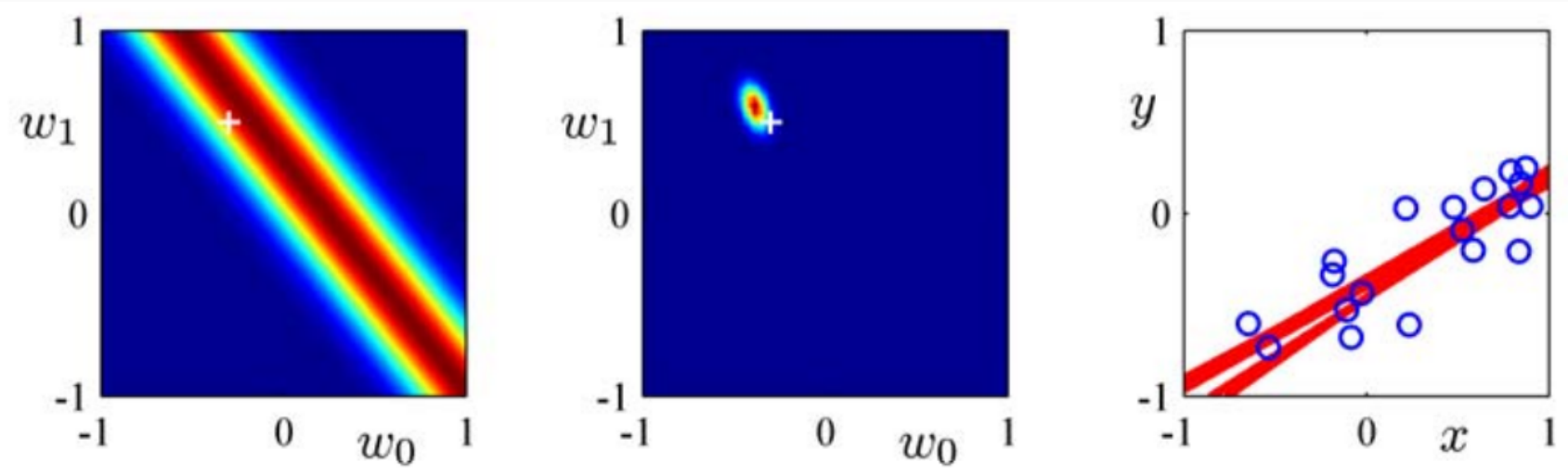
Пример



Пример



Пример



Обобщение

- Можно слегка обобщить – рассмотреть априорное распределение более общего вида

$$p(\mathbf{w} \mid \alpha) = \left[\frac{q}{2} \left(\frac{\alpha}{2} \right)^{1/q} \frac{1}{\Gamma(1/q)} \right]^M e^{-\frac{\alpha}{2} \sum_{j=1}^{M} |w_j|^q}.$$

Упражнение. Подсчитайте логарифм правдоподобия.

Лассо

- 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»?

Лассо

- Мы можем переписать регрессию с регуляризатором по-другому:

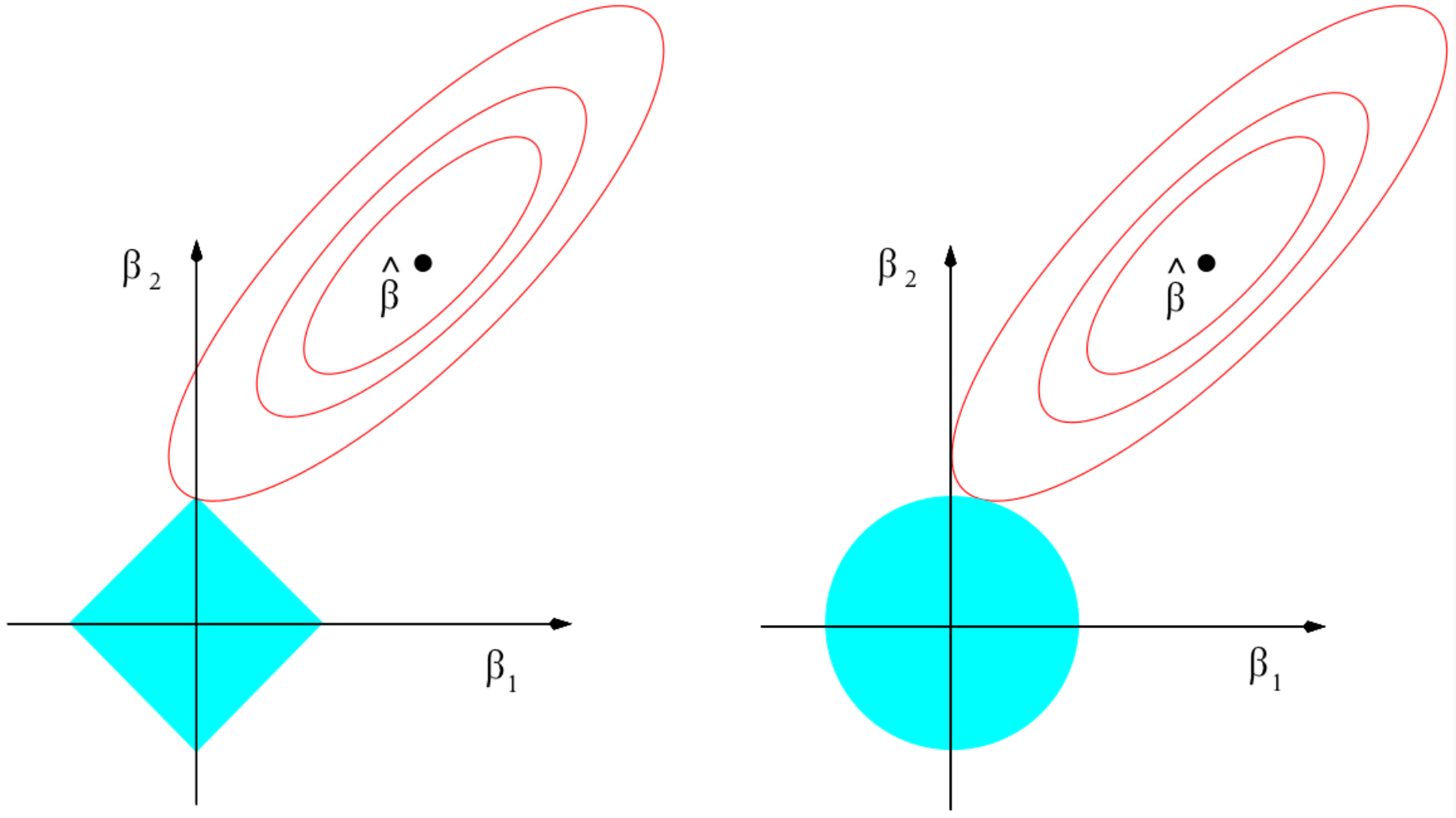
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \left\{ \frac{1}{2} \sum_{i=1}^{N} (f(x_i, \mathbf{w}) - y_i)^2 + \lambda \sum_{j=0}^{p} |w_j| \right\},$$

эквивалентно

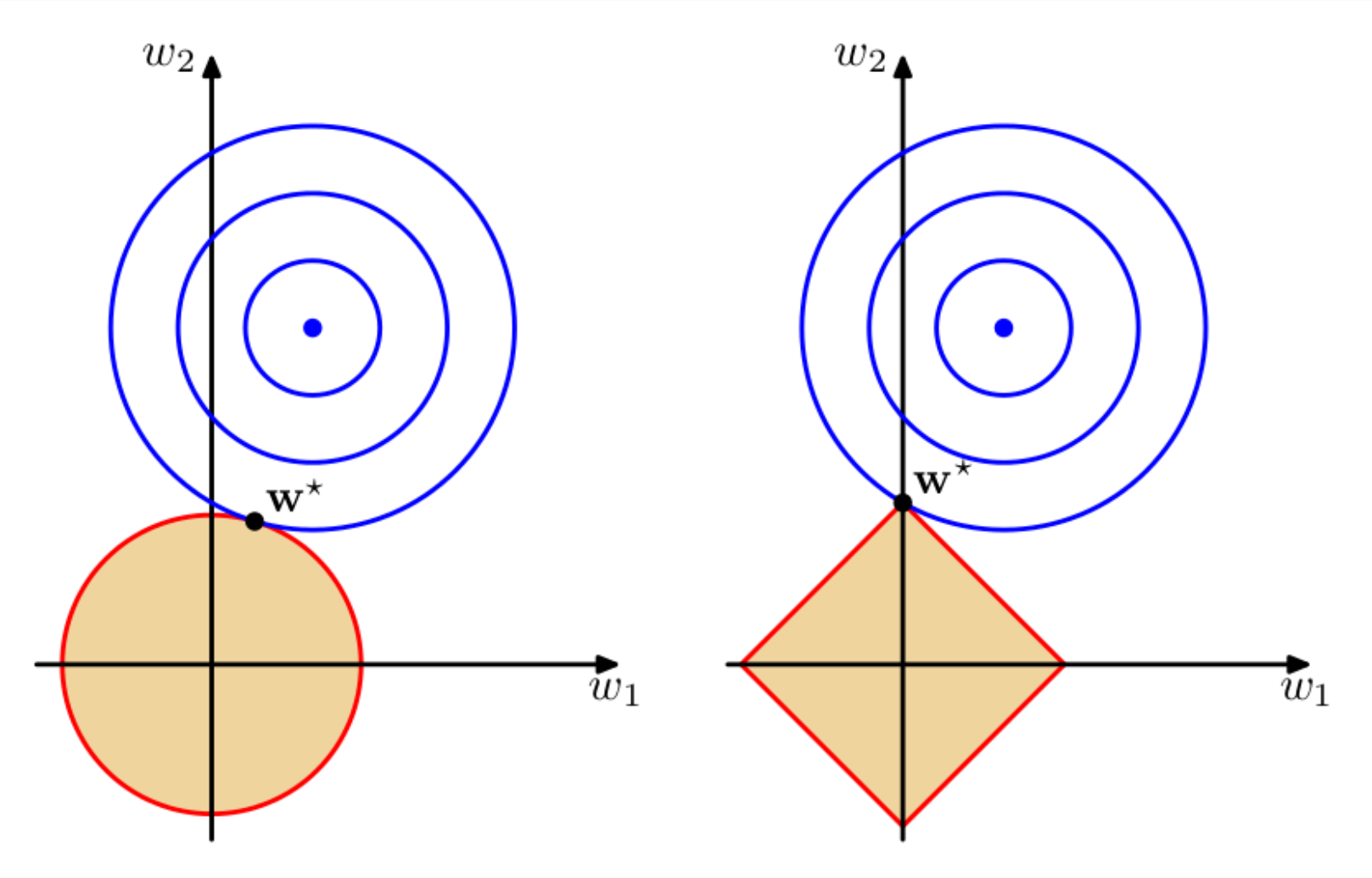
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \left\{ \frac{1}{2} \sum_{i=1}^{N} (f(x_i, \mathbf{w}) - y_i)^2 \right\} \text{ при } \sum_{j=0}^{p} |w_j| \leq t.$$

Упражнение. Докажите это.

Гребень и лассо



Гребень и лассо



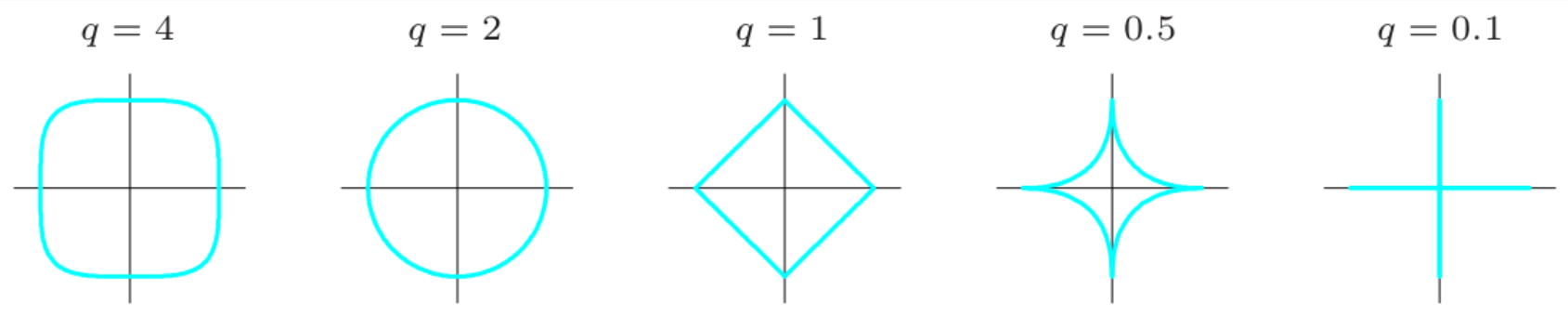
Обобщение

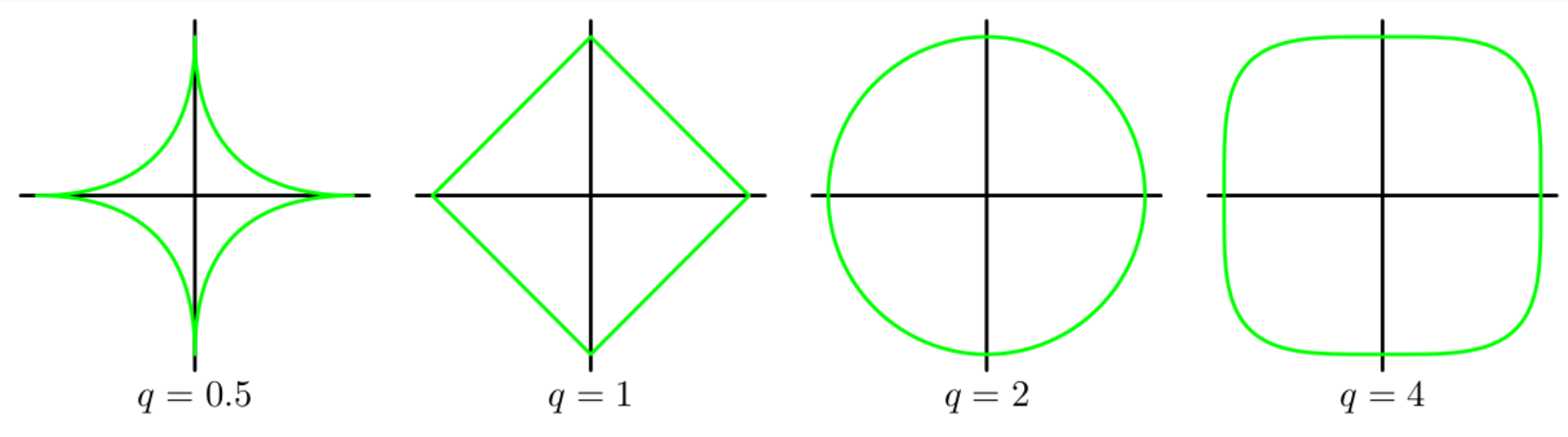
- Можно рассмотреть обобщение гребневой и лассо-регрессии:

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

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

Разные q





Предсказания в линейной регрессии

Предсказание в линейной регрессии

- Теперь давайте вернёмся к байесовской постановке:
  1. найти апостериорное распределение на гипотезах/параметрах:

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

     (и/или найти максимальную апостериорную гипотезу $\arg\max_\theta p(\theta \mid D)$);
  2. найти апостериорное распределение исходов дальнейших экспериментов:

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

Предсказание в линейной регрессии

- В прошлый раз мы нашли апостериорное распределение: для гауссовского априорного

$$p(\mathbf{w} \mid \alpha) = N(\mathbf{w} \mid 0, \frac{1}{\alpha}\mathbf{I})$$

мы нашли

$$\begin{aligned} p(\mathbf{w} \mid \mathbf{t}, \alpha, \beta) &= N(\mathbf{w} \mid \mu_N, \Sigma_N), \\ \mu_N &= \Sigma_N \left(\Sigma_0^{-1}\mu_0 + \beta\Phi^\top \mathbf{t}\right), \\ \Sigma_N &= \left(\Sigma_0^{-1} + \beta\Phi^\top\Phi\right)^{-1}, \end{aligned}$$

где $\beta = \frac{1}{\sigma^2}$ (precision нормального распределения).

Предсказание в линейной регрессии

- Теперь сделаем следующий шаг – найдём апостериорное распределение наших предсказаний

$$p(t \mid \mathbf{t}, \alpha, \beta) = \int p(t \mid \mathbf{w}, \beta) p(\mathbf{w} \mid \mathbf{t}, \alpha, \beta) d\mathbf{w}.$$

- Это свёртка двух гауссианов, и получается...

Предсказание в линейной регрессии

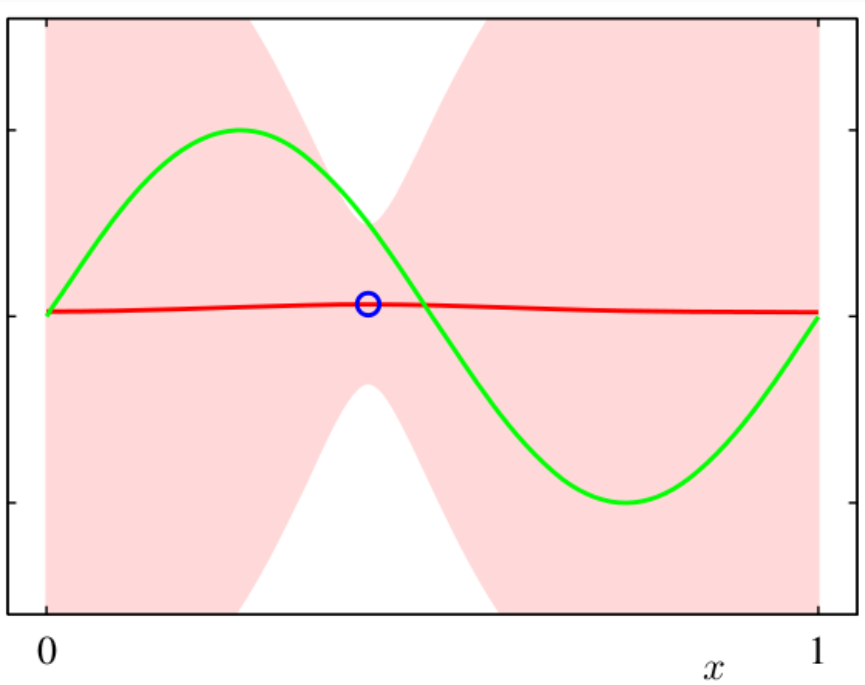
- ...тоже гауссиан:

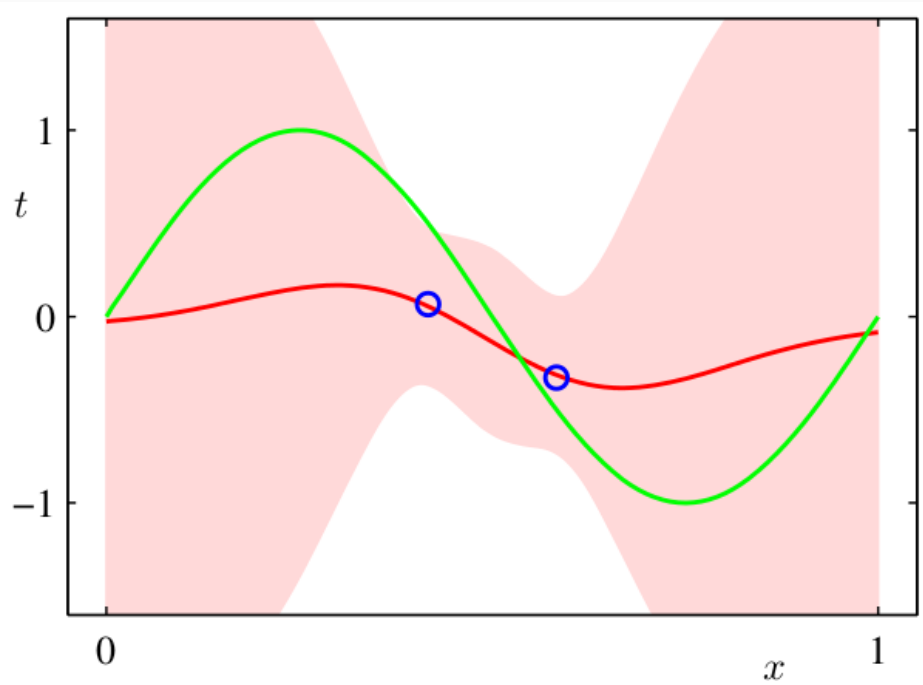
$$p(t \mid \mathbf{t}, \alpha, \beta) = N(t \mid \mu_N^\top \phi(\mathbf{x}), \sigma_N^2),$$
$$\text{где } \sigma_N^2 = \frac{1}{\beta} + \phi(\mathbf{x})^\top \Sigma_N \phi(\mathbf{x}).$$

- Т.е. дисперсия складывается из шума в данных β и дисперсии параметров $\mathbf{w}$; гауссианы независимы, и их дисперсии просто складываются.

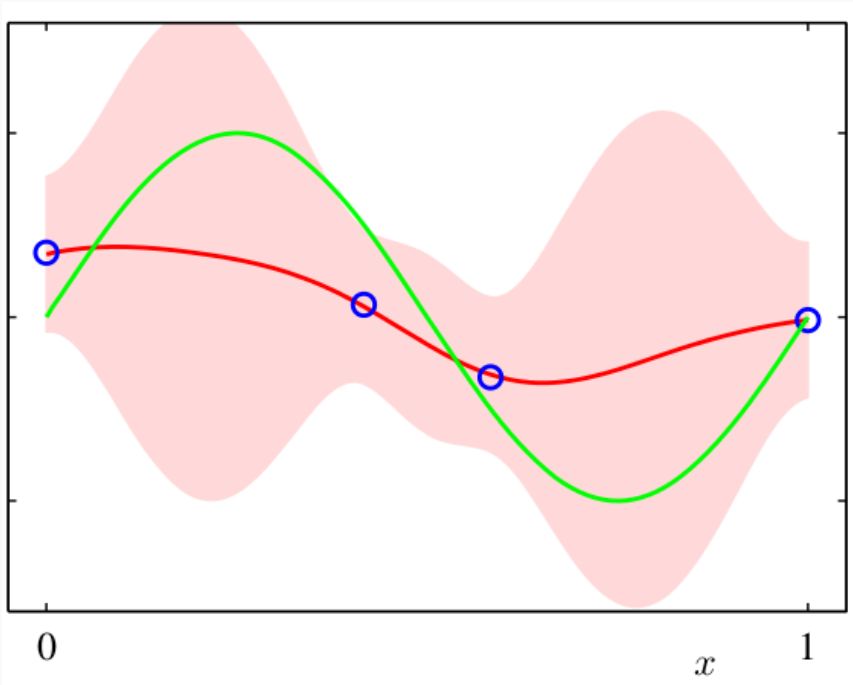
Упражнение. Оценка всё время уточняется: $\sigma_{N+1}^2 \leq \sigma_N^2$.

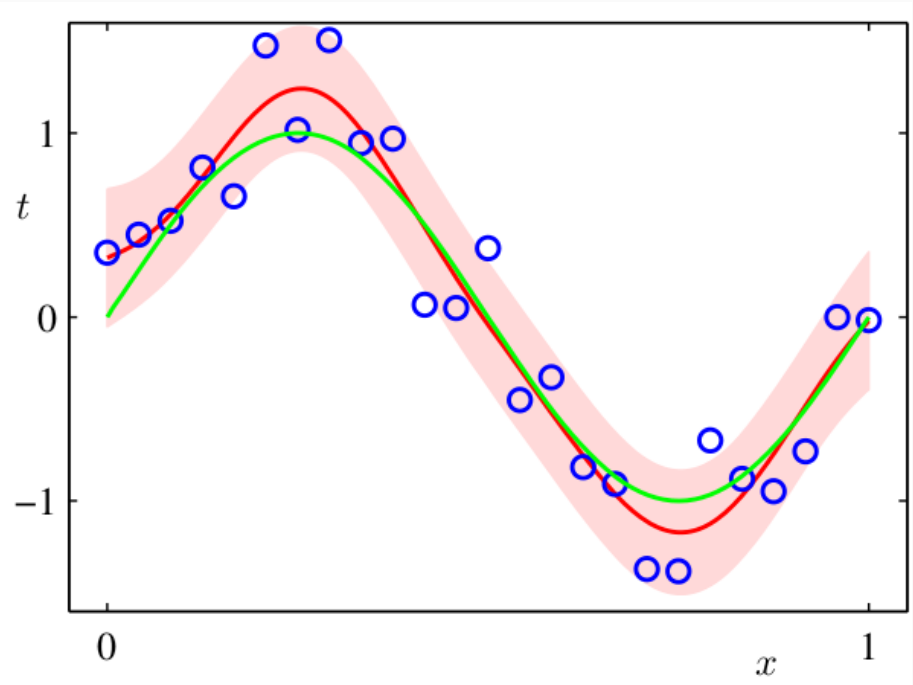
Предсказания



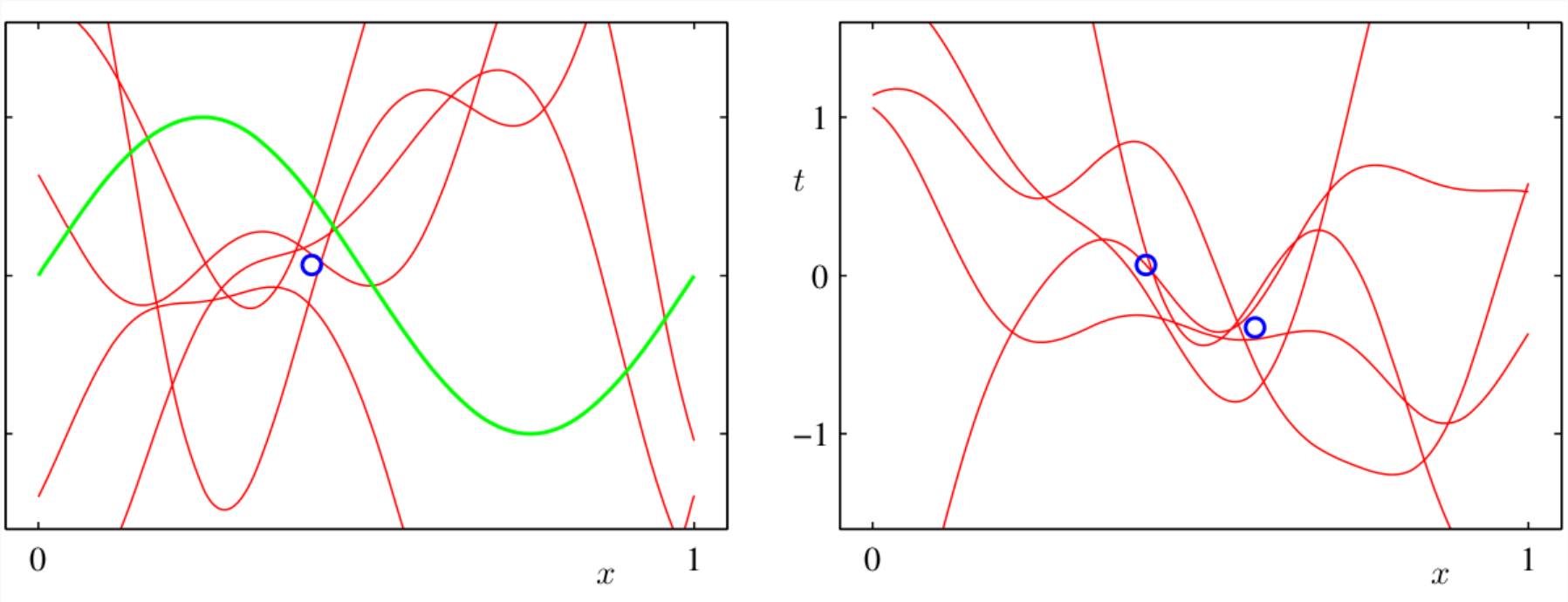


Предсказания

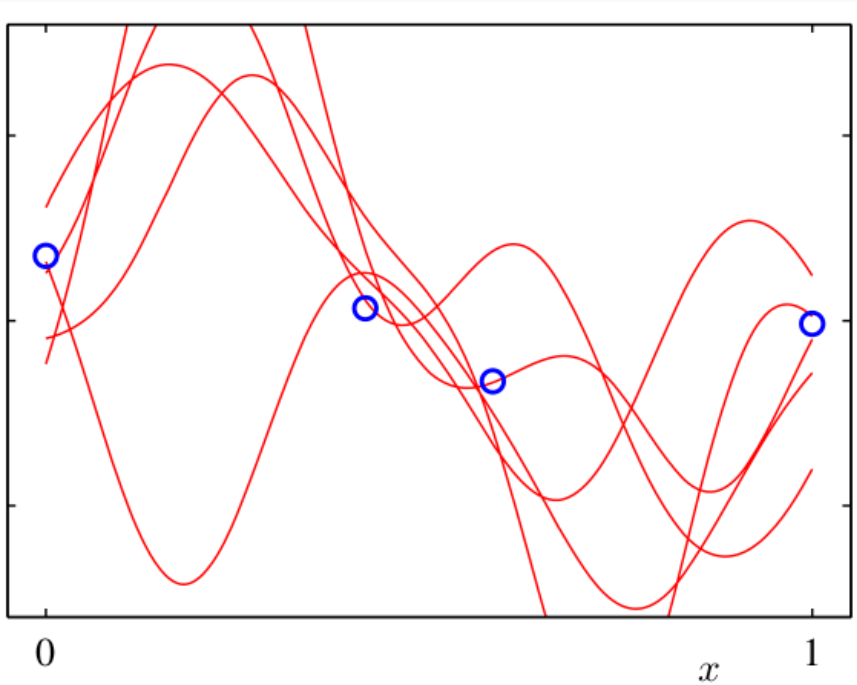


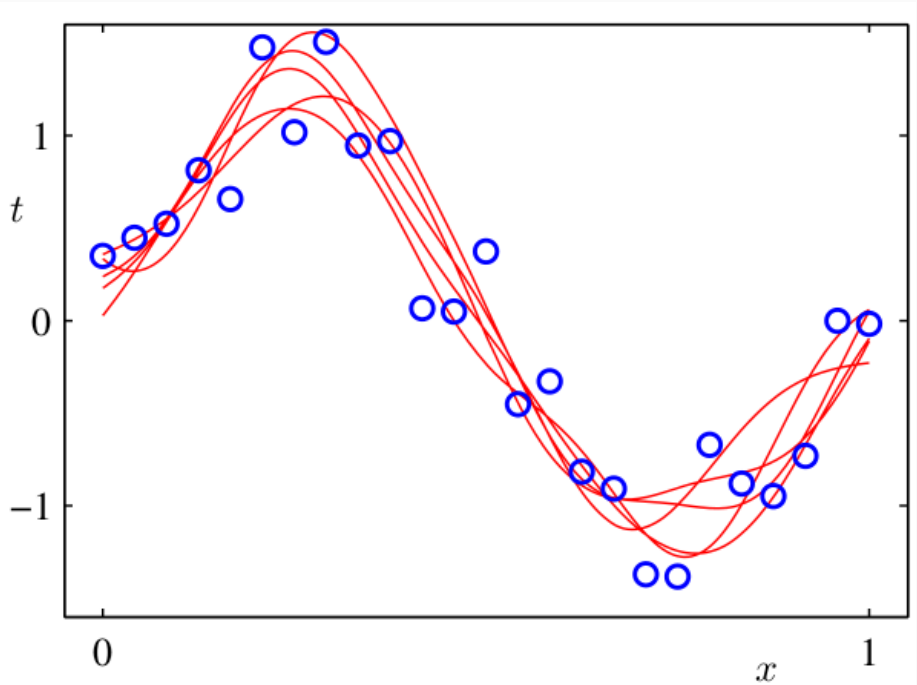


Предсказания



Предсказания





Спасибо!

Спасибо за внимание!