ЛИНЕЙНАЯ РЕГРЕССИЯ ПО-БАЙЕСОВСКИ

Сергей Николенко

СПбГУ — Санкт-Петербург
16 сентября 2020 г.

Random facts:

- 16 сентября 1620 г. первые английские пилигримы отплыли из Плимута в Новый свет на корабле «Мэйфлауэр»
- 16 сентября 1857 г. Джейн Пирпонт из Бостона получила авторские права на песню «One Horse Open Sleigh», более известную как «Jingle Bells»
- 16 сентября 1976 г. многократный рекордсмен и чемпион мира по подводному плаванию Шаварш Карапетян спас с 10-метровой глубины 20 человек, когда переполненный троллейбус упал с моста в Ереванское водохранилище
- 16 сентября 1992 г. — Чёрная среда для британской валюты: фунт стерлингов резко подешевел, потеряв в итоге до конца года более 25% по отношению к доллару; много говорили о том, что это проделки Джорджа Сороса, но это, кажется, не так
- 16 сентября 1996 г. в США был арестован, наверное, самый знаменитый хакер в истории, Кевин Митник
- 16 сентября 1998 г. в Париже прошла премьера мюзикла «Notre-Dame de Paris»

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:

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

Байесовская регуляризация

- Тогда наша задача – посчитать

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

- 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:

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

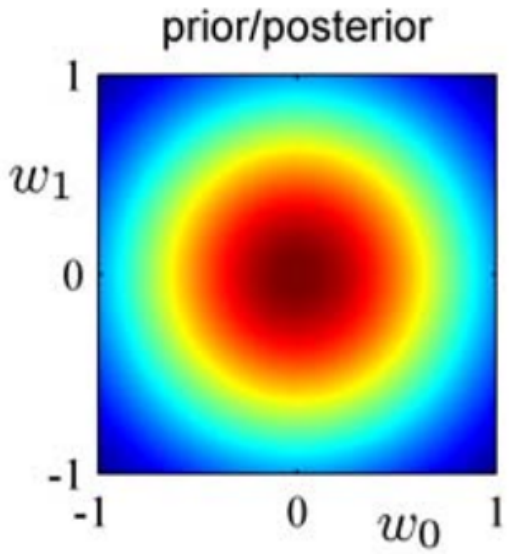
то логарифм правдоподобия получится

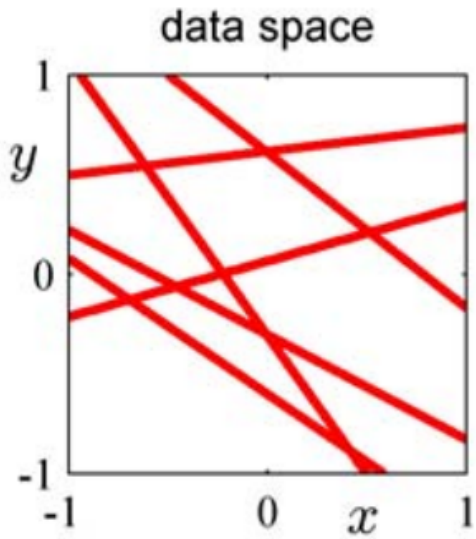
$$\ln p(\mathbf{w} \mid \mathbf{t}) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(t_n - \mathbf{w}^\top\phi(\mathbf{x}_n)\right)^2 - \frac{\alpha}{2}\mathbf{w}^\top\mathbf{w} + \text{const},$$

то есть в точности гребневая регрессия.

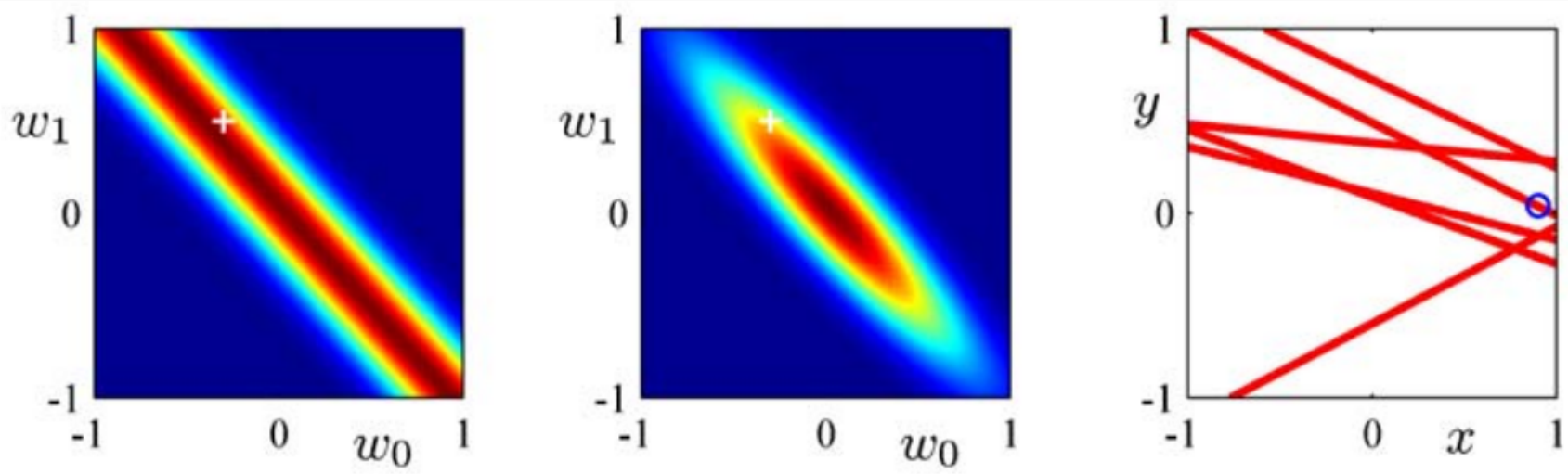
Пример

likelihood

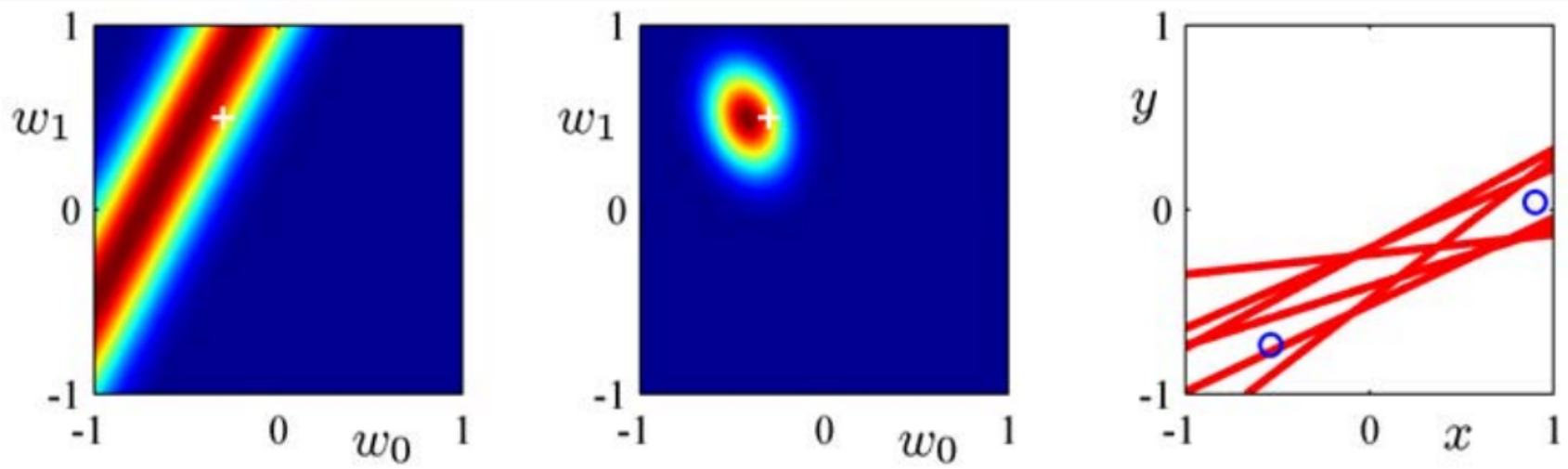




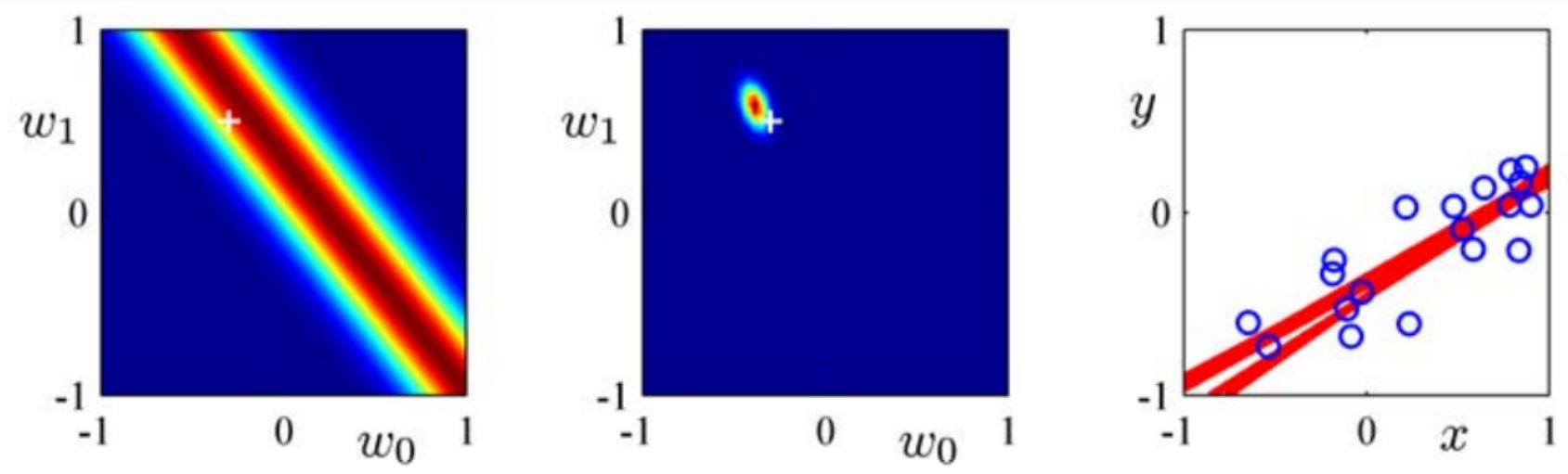
Пример



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.

И снова о регуляризации

Ещё раз о регуляризации

- 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 $\mathbf{w}$.

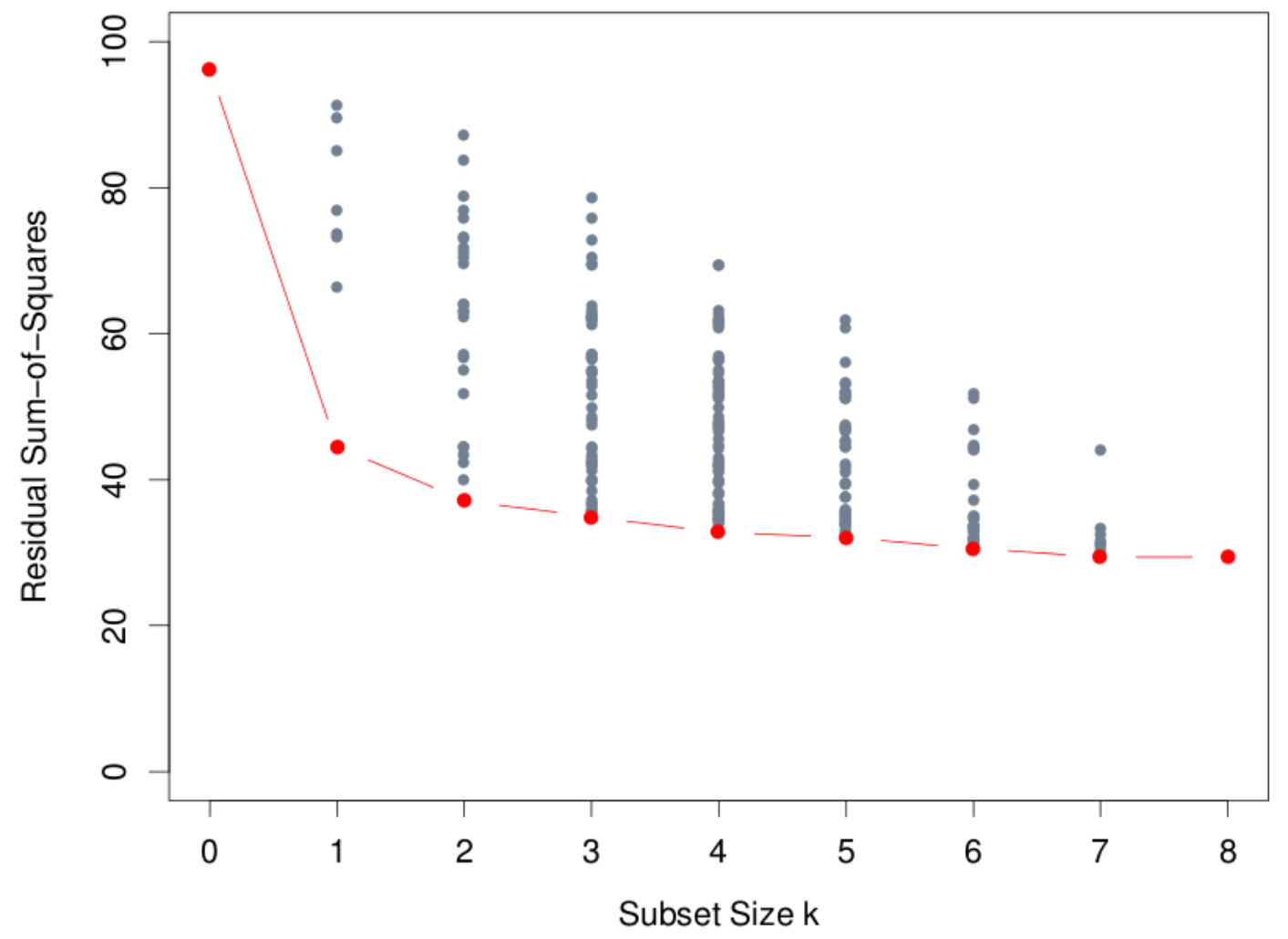
Subset selection

- 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.

Subset selection

- 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



SVD и гребневая регрессия

- Запишем SVD (разложение $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$) для матрицы $\mathbf{X}$ в линейной регрессии сначала с обычным least squares; посмотрим на предсказание $\hat{\mathbf{y}} = \mathbf{X}\mathbf{w}$:

$$\mathbf{X}\mathbf{w}^{\text{ls}} = \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{y} = \mathbf{U}\mathbf{U}^\top\mathbf{y},$$

где, кстати, $\mathbf{U}^\top\mathbf{y}$ – просто координаты $\mathbf{y}$ в базисе $\mathbf{U}$.

SVD и гребневая регрессия

- А теперь запишем SVD для гребневой регрессии:

$$\mathbf{X}\mathbf{w}^{\text{ridge}} = \mathbf{X}\left(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^\top\mathbf{y} =$$

$$\mathbf{U}\mathbf{D}\left(\mathbf{D}^2 + \lambda\mathbf{I}\right)^{-1}\mathbf{D}\mathbf{U}^\top\mathbf{y} = \sum_{j=1}^{p}\mathbf{u}_j\frac{d_j^2}{d_j^2+\lambda}\mathbf{u}_j^\top\mathbf{y},$$

где $\mathbf{u}_j$ – столбцы $\mathbf{U}$.

SVD и гребневая регрессия

- $\mathbf{X}\mathbf{w}^{\text{ridge}} = \sum_{j=1}^{p} \mathbf{u}_j \frac{d_j^2}{d_j^2+\lambda} \mathbf{u}_j^\top \mathbf{y}$.
- Иначе говоря, гребневая регрессия *стягивает* координаты вектора $\mathbf{y}$ в базисе $\mathbf{U}$ (т.е. $\mathbf{u}_j^\top \mathbf{y}$), причём сильнее всего уменьшаются самые *маленькие* координаты.
- Гребневая регрессия больше всего стягивает $\mathbf{X}$ по тем направлениям, где дисперсия $\mathbf{X}$ минимальна.

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:

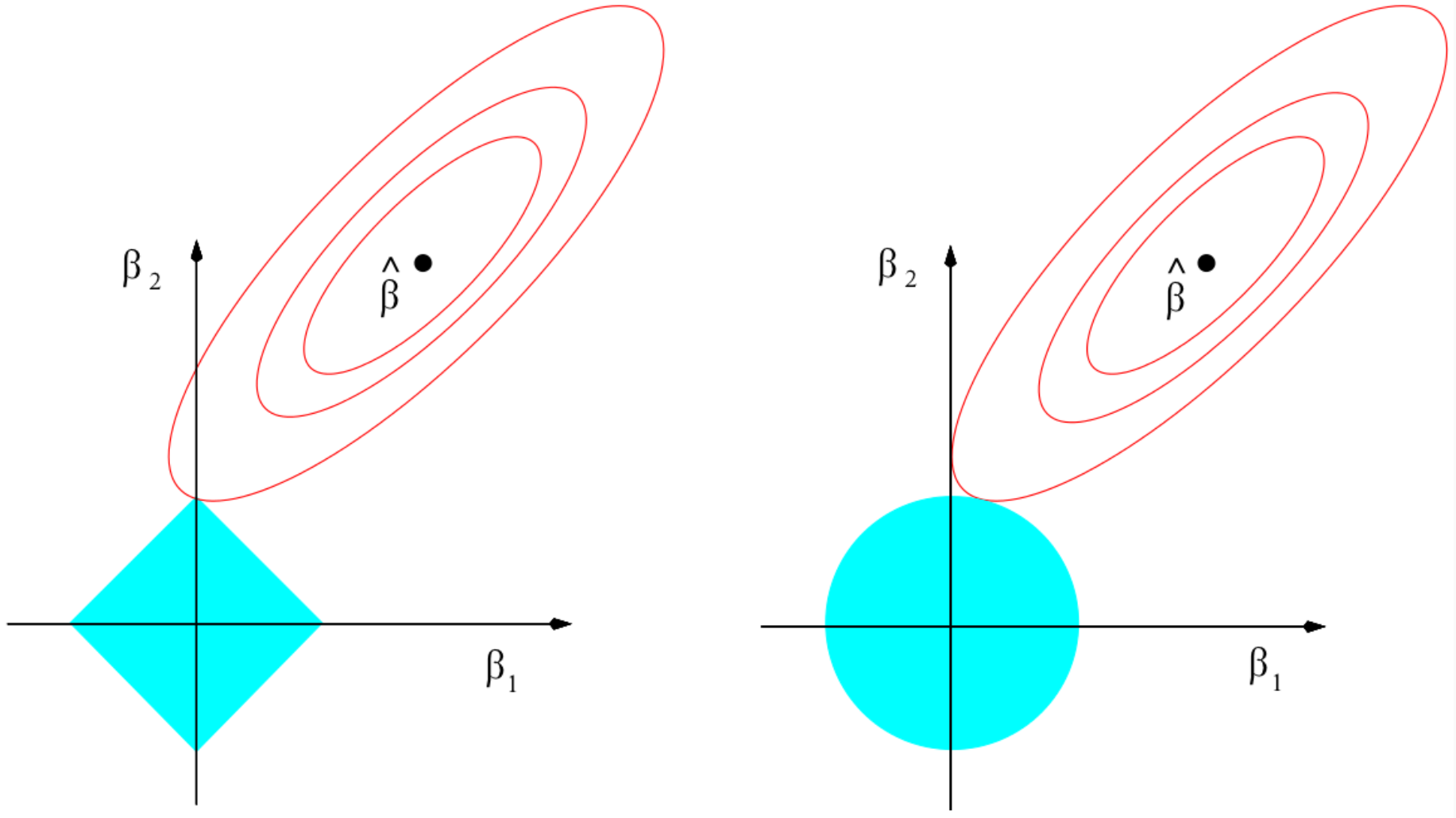
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \left\{ \frac{1}{2} \sum_{i=1}^{N} (f(x_i, \mathbf{w}) - y_i)^2 + \lambda \sum_{j=0}^{p} |w_j| \right\},$$

эквивалентно

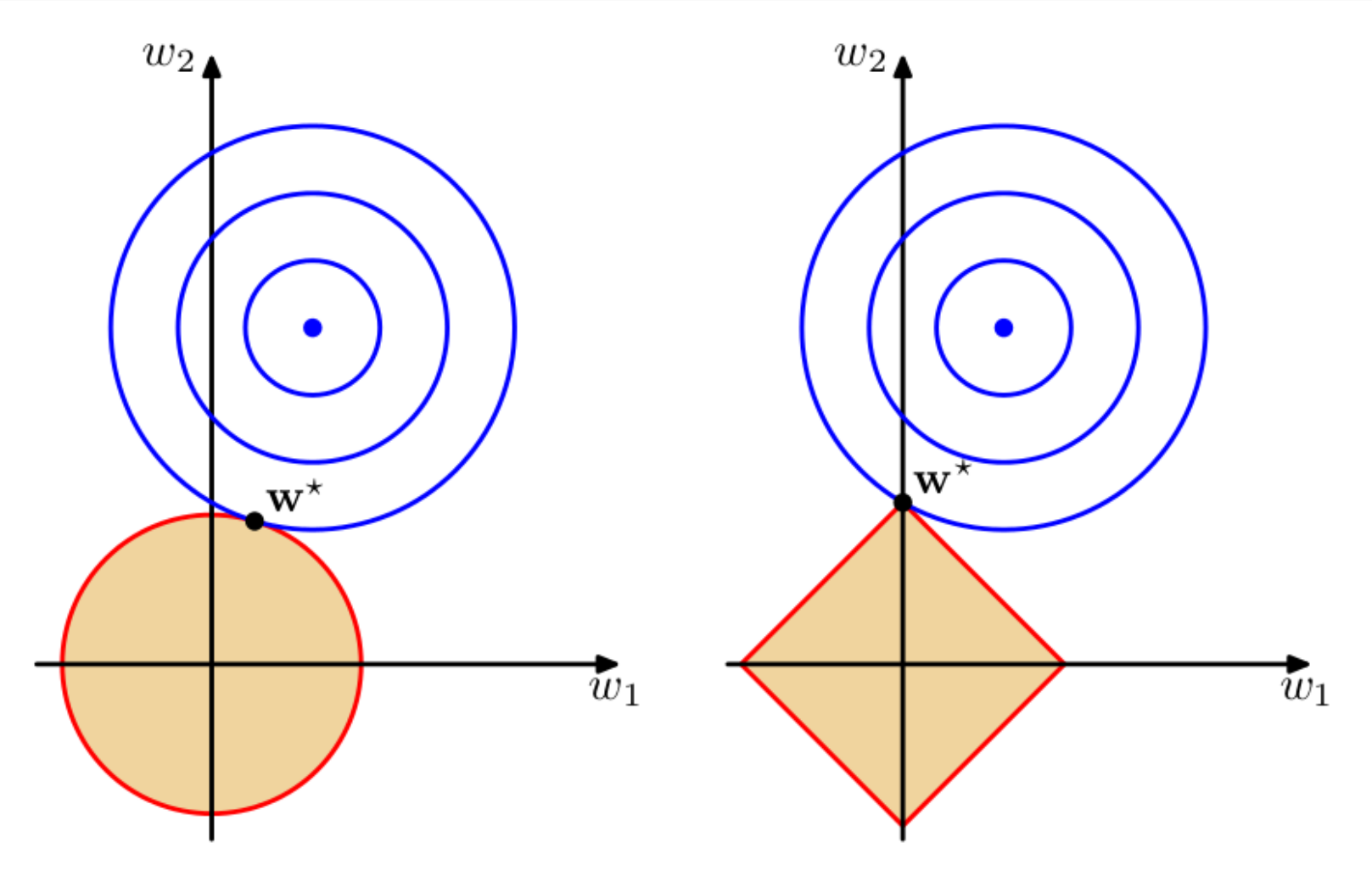
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \left\{ \frac{1}{2} \sum_{i=1}^{N} (f(x_i, \mathbf{w}) - y_i)^2 \right\} \text{ при } \sum_{j=0}^{p} |w_j| \leq t.$$

Упражнение. Докажите это.

Гребень и лассо



Гребень и лассо



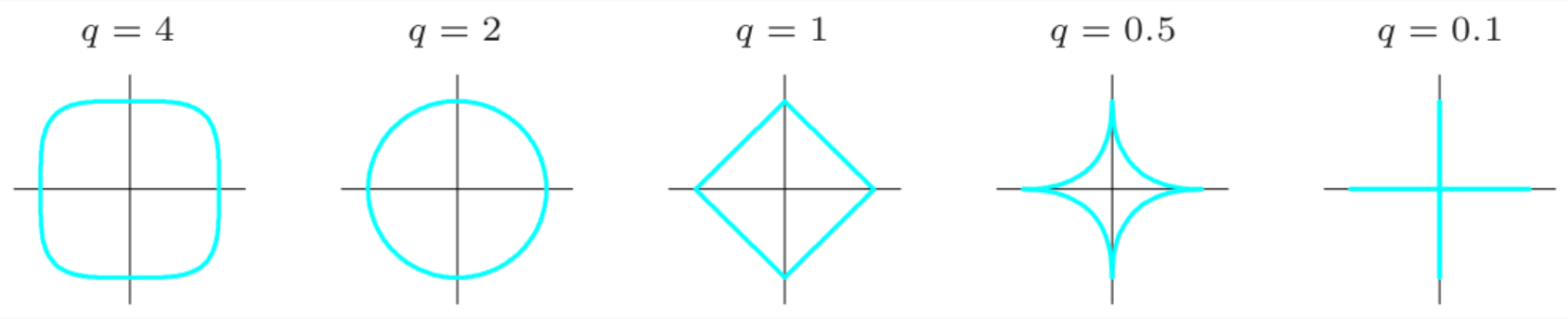
Обобщение

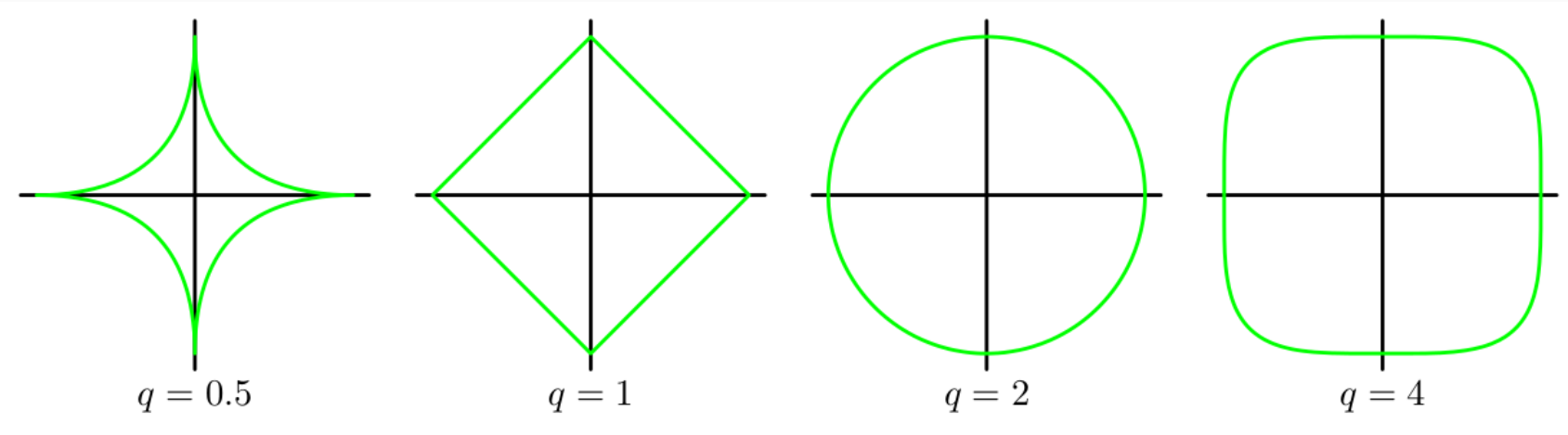
- Можно рассмотреть обобщение гребневой и лассо-регрессии:

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

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

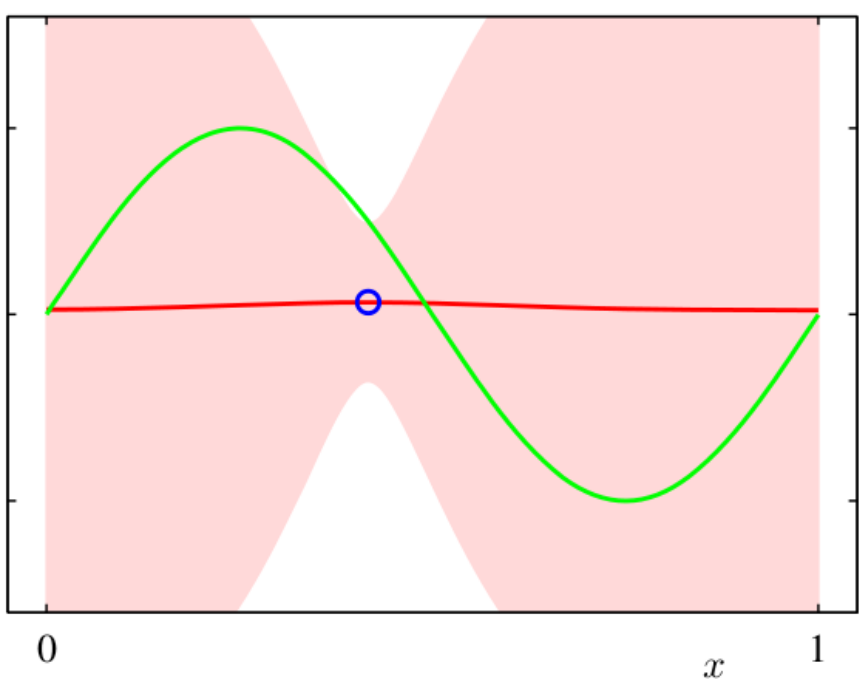
- ...тоже гауссиан:

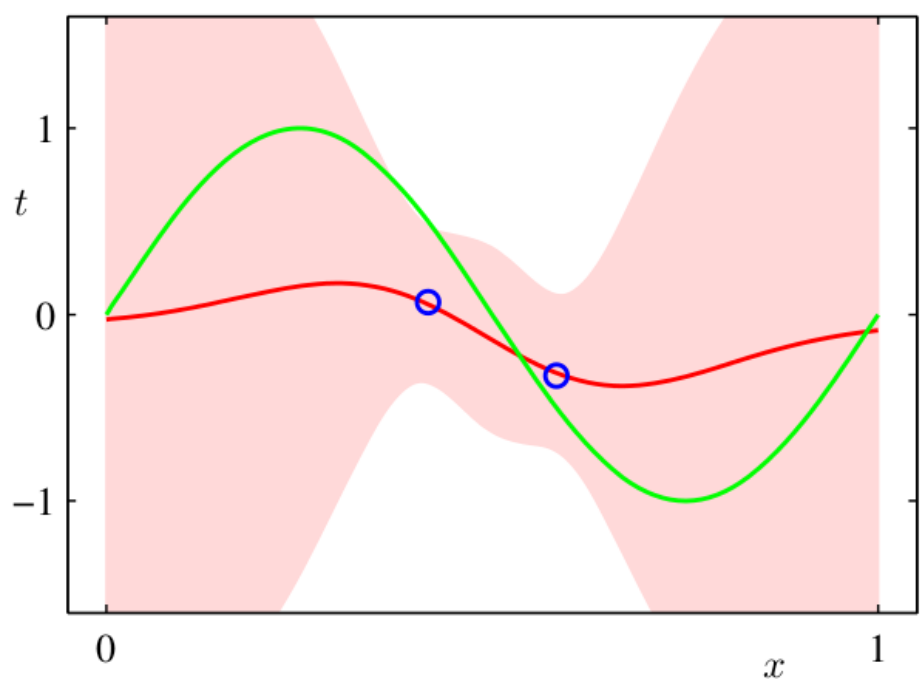
$$p(t \mid \mathbf{t}, \alpha, \beta) = N(t \mid \mu_N^\top \phi(\mathbf{x}), \sigma_N^2),$$
$$\text{где } \sigma_N^2 = \frac{1}{\beta} + \phi(\mathbf{x})^\top \Sigma_N \phi(\mathbf{x}).$$

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

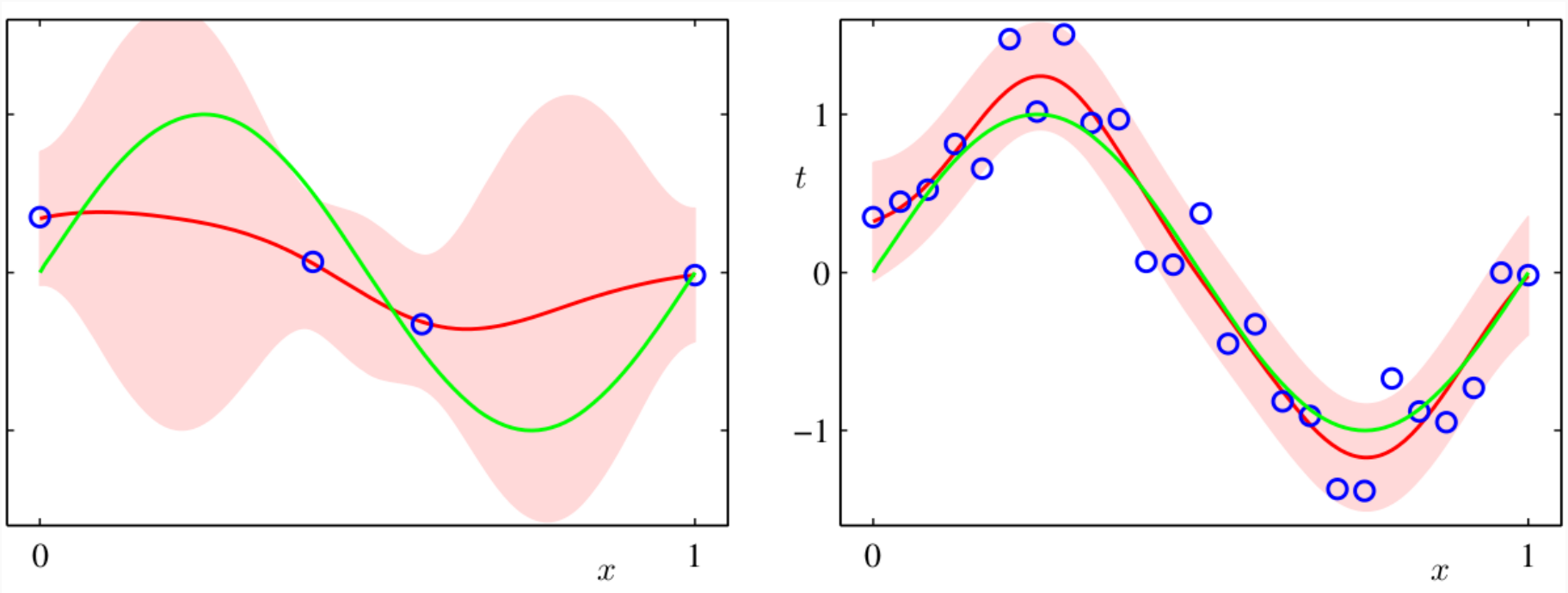
Упражнение. Оценка всё время уточняется: $\sigma_{N+1}^2 \leq \sigma_N^2$.

Предсказания

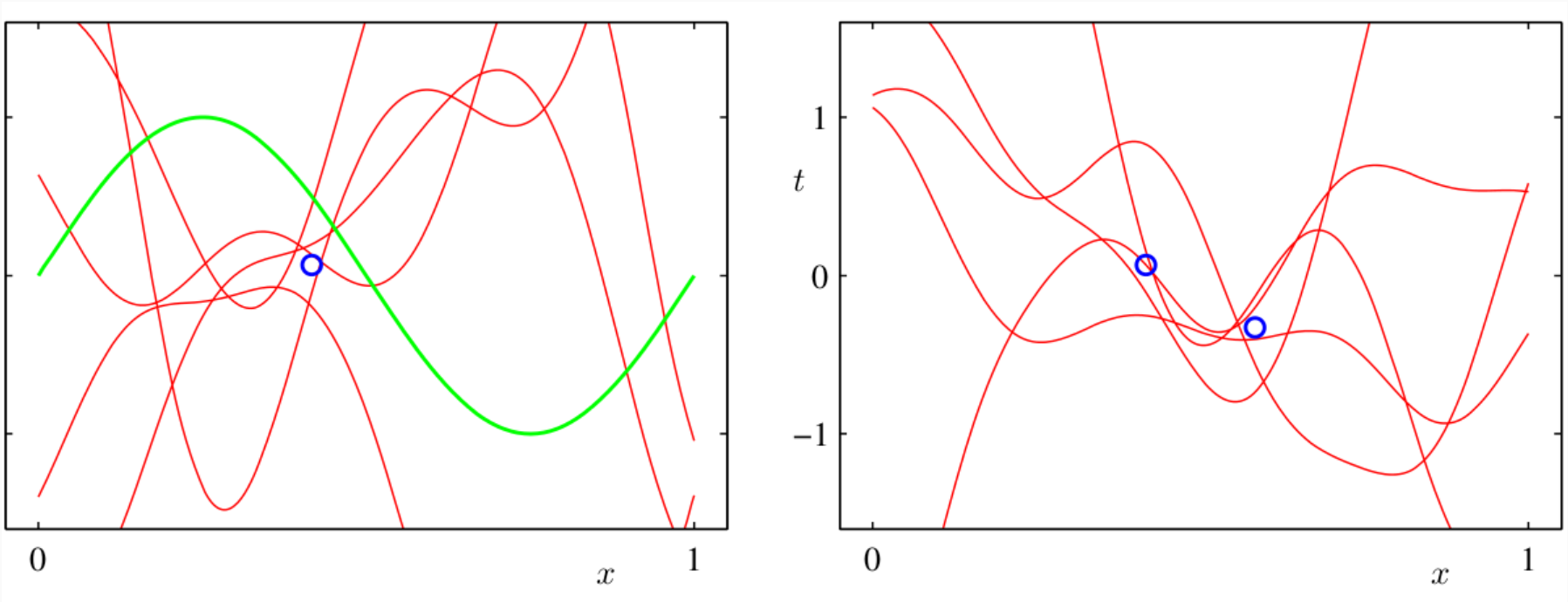




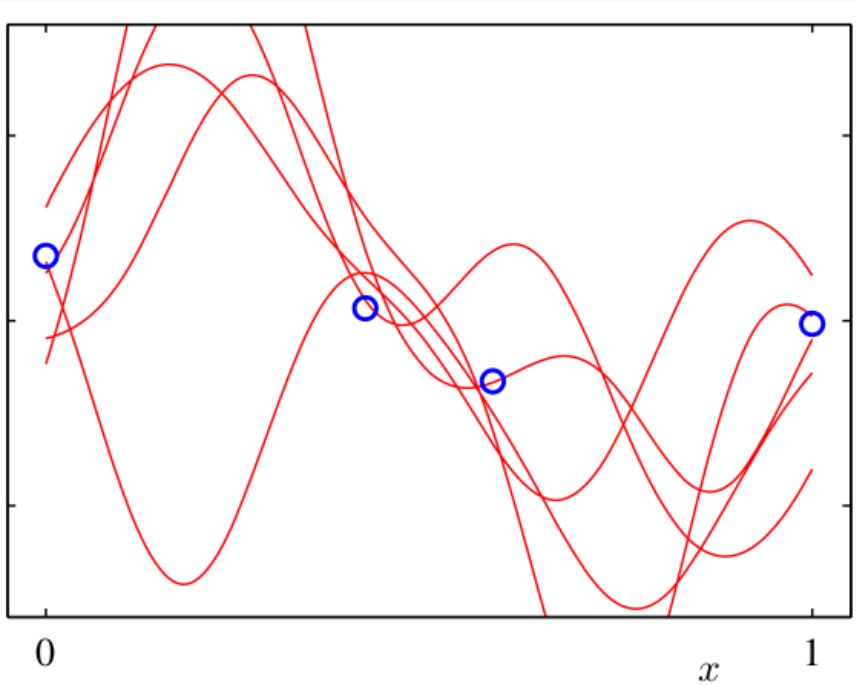
Предсказания

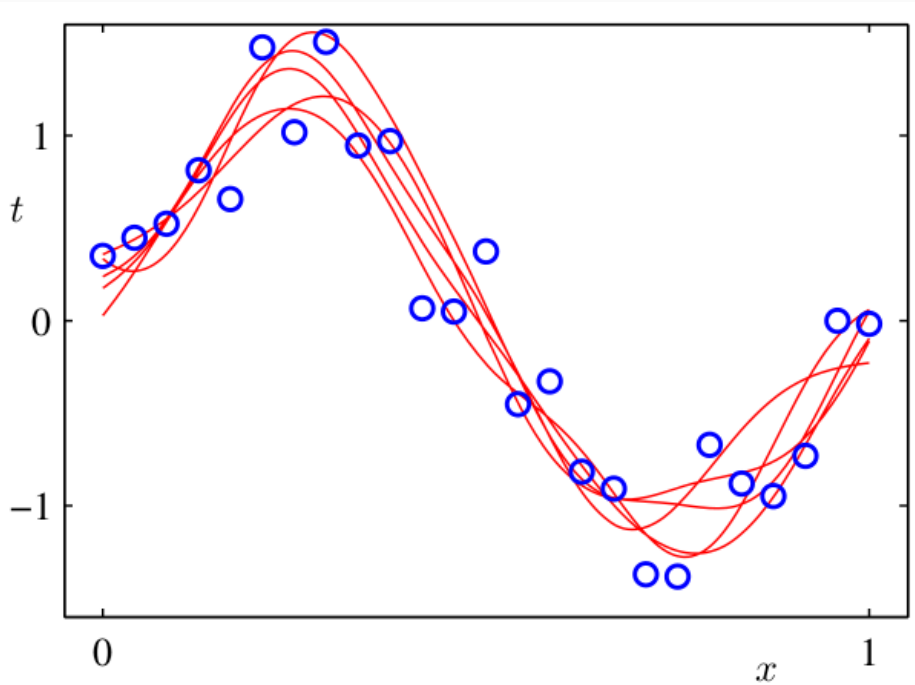


Предсказания



Предсказания





Байесовский вывод для гауссиана

Нормальное распределение: фиксируем σ

- На самом деле всё это — байесовский вывод для нормального распределения:

$$p(x_1, \dots, x_n \mid \mu, \sigma^2) \propto \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2\right).$$

- Хотим: найти сопряжённое априорное распределение, подсчитать правдоподобие, решить задачу предсказания.
- Для начала зафиксируем σ^2 и будем в качестве параметра рассматривать только μ.

Нормальное распределение: фиксируем σ

- Сопряжённое априорное распределение для μ при фиксированном σ^2 тоже нормальное и выглядит как

$$p(\mu \mid \mu_0, \sigma_0^2) \propto \frac{1}{\sigma_0^n} \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right).$$

- Обычно выбирают $\mu_0 = 0$, $\sigma_0^2 \to \infty$ (порой буквально).
- Давайте рассмотрим сначала случай ровно одного наблюдения x и найдём $p(\mu \mid x)$.

Нормальное распределение: фиксируем σ

- При нашем априорном распределении у μ и x совместное нормальное распределение:

$$x = \mu + \sigma\epsilon, \quad \mu = \mu_0 + \sigma_0\delta, \quad \epsilon, \delta \sim \mathcal{N}(0, 1).$$

Упражнение. Пусть (z_1, z_2) – случайные величины с совместным нормальным распределением. Докажите, что случайная величина $z_1 \mid z_2$ распределена нормально с параметрами

$$E(z_1 \mid z_2) = E(z_1) + \frac{Cov(z_1, z_2)}{Var(z_2)}\left(z_2 - E(z_2)\right),$$

$$Var(z_1 \mid z_2) = Var(z_1) - \frac{Cov^2(z_1, z_2)}{Var(z_2)}$$

$(Var(x) = E\left[(x - Ex)^2\right], Cov(x, y) = E\left[(x - Ex)(y - Ey)\right])$.

Нормальное распределение: фиксируем σ

- В нашем случае:

$$
\begin{aligned}
x = \mu + \sigma\epsilon, \quad \mu = \mu_0 + \sigma_0\delta, \quad \epsilon, \delta \sim \mathcal{N}(0,1),\\
E(x) = \mu_0,\\
Var(x) = E(Var(x \mid \mu)) + Var(E(x \mid \mu)) = \sigma^2 + \sigma_0^2,\\
Cov(x, \mu) = E\left[(x - \mu_0)(\mu - \mu_0)\right] = \sigma_0^2.
\end{aligned}
$$

- Применив упражнение, получаем:

$$
\begin{aligned}
E(\mu \mid x) = \mu_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}(x - \mu_0) = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}x + \frac{\sigma^2}{\sigma_0^2 + \sigma^2}\mu_0,\\
Var(\mu \mid x) = \frac{\sigma^2\sigma_0^2}{\sigma_0^2 + \sigma^2} = \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}}.
\end{aligned}
$$

Нормальное распределение: фиксируем σ

- Итого:

$$p(\mu \mid x) \sim \mathcal{N}\left(\frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} x + \frac{\sigma^2}{\sigma_0^2 + \sigma^2}\mu_0, \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}\right)^{-1}\right).$$

- Опять же, сложные вычисления можно забыть и пользоваться этими формулами.
- Замечание: часто используют $\tau = \frac{1}{\sigma^2}$ как параметр нормального распределения (precision). Тогда

$$\tau_{\mu|x} = \tau_\mu + \tau.$$

Нормальное распределение: фиксируем σ

- А что, если данных больше, $x_1, \dots, x_n$?
- Тогда можно повторить всё то же самое, а можно заметить, что набор данных описывается своим средним.

Упражнение. Докажите, что если $p(x_i \mid \mu) \sim \mathcal{N}(\mu, \sigma^2)$ и x_i независимы, то $p(\bar{x} \mid \mu) \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Нормальное распределение: фиксируем σ

- Для апостериорной вероятности будет

$$p(\mu \mid x_1, \dots, x_n) \propto p(x_1, \dots, x_n \mid \mu) p(\mu) \propto p(\bar{x} \mid \mu) p(\mu) \propto p(\mu \mid \bar{x}).$$

- Подставляя в наш предыдущий результат, получим:

$$p(\mu \mid x_1, \dots, x_n) \sim \mathcal{N}\left(\frac{\sigma_0^2}{\sigma_0^2 + \frac{\sigma^2}{n}} x + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \mu_0, \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1} \right).$$

Нормальное распределение: фиксируем μ

- Если зафиксировать μ и менять σ^2, то сопряжённым априорным распределением будет обратное гамма-распределение:

$$p(\sigma^2 \mid \alpha, \beta) \propto IG(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} \exp\left(\frac{-\beta}{z}\right).$$

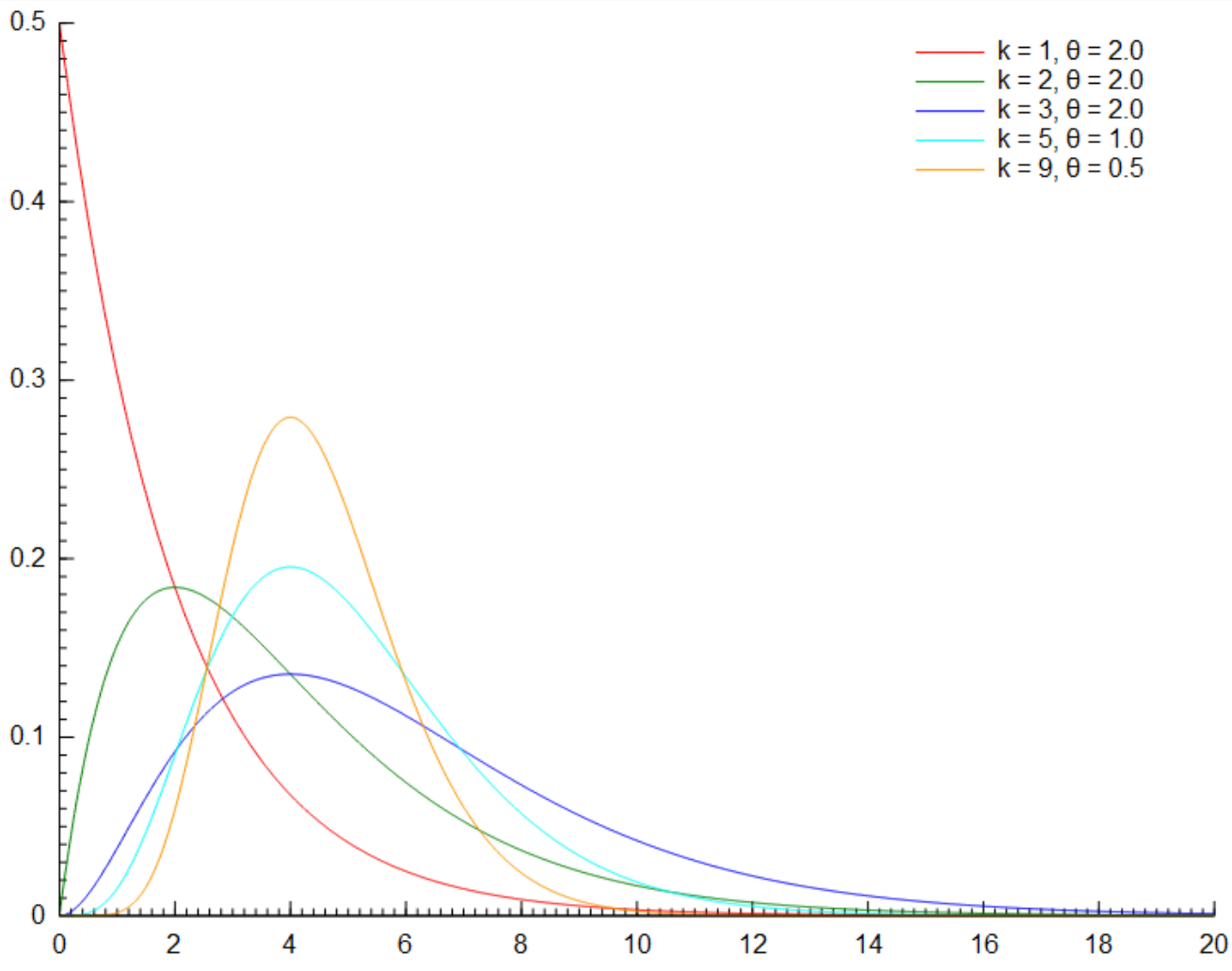
- Тогда в апостериорном распределении будет

$$p(\sigma^2 \mid x_1, \dots, x_n, \alpha, \beta) \propto IG\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum(x_i - \mu)\right).$$

- А в терминах $\tau = \frac{1}{\sigma^2}$ будет обычное гамма-распределение:

$$p(\tau \mid x_1, \dots, x_n, \alpha, \beta) \propto \text{Gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum(x_i - \mu)\right).$$

Гамма--распределение



Когда и μ, и σ^2 меняются

- Что делать, когда и μ, и σ^2 меняются?
- Можно было бы предположить, что μ и σ^2 независимы; тогда просто априорное распределение будет

$$p(\mu, \sigma \mid \mu_0, \sigma_0, \alpha, \beta) \propto \mathcal{N}(\mu_0, \sigma_0^2) \cdot IG(\alpha, \beta).$$

- К сожалению, это распределение не будет сопряжённым к нормальному. Почему?

Когда и μ, и σ^2 меняются

- Что делать, когда и μ, и σ^2 меняются?
- Можно было бы предположить, что μ и σ^2 независимы; тогда просто априорное распределение будет

$$p(\mu, \sigma \mid \mu_0, \sigma_0, \alpha, \beta) \propto \mathcal{N}(\mu_0, \sigma_0^2) \cdot IG(\alpha, \beta).$$

- К сожалению, это распределение не будет сопряжённым к нормальному. Почему?
- Потому что μ и σ^2 зависимы. :) Новая точка x вводит зависимость между ними.
- В результате получается распределение Стьюдента.

Экспоненциальное семейство

- Вообще говоря, всё, о чём мы говорили – частные случаи *экспоненциального семейства* распределений:

$$p(\mathbf{x} \mid \eta) = h(\mathbf{x})g(\eta)e^{\eta^\top \mathbf{u}(\mathbf{x})}.$$

- η называются *естественными параметрами* (natural parameters).

Экспоненциальное семейство

- Например, распределение Бернулли:

$$p(x \mid \mu) = \mu^x (1-\mu)^{1-x} = e^{x \ln \mu + (1-x)\ln(1-\mu)} = \\ = (1-\mu) e^{\ln\left(\frac{\mu}{1-\mu}\right) x},$$

и естественный параметр получился $\eta = \ln\left(\frac{\mu}{1-\mu}\right)$:

$$p(x \mid \eta) = \sigma(-\eta) e^{-\eta x},$$

где $\sigma(y) = \frac{1}{1+e^{-y}}$ – *сигмоид-функция*.

Экспоненциальное семейство

- Для мультиномиального распределения с параметрами $\mu_1, \ldots, \mu_{M-1}$ получаются

$$\eta_k = \ln\left(\frac{\mu_k}{1 - \sum_j \mu_j}\right) \quad \text{и}$$

$$p(\mathbf{x} \mid \eta) = \left(1 + \sum_{k=1}^{M-1} e^{\eta_k}\right)^{-1} e^{\eta^\top \mathbf{x}}.$$

Упражнение. Проверьте!

Экспоненциальное семейство

- Так вот, для распределений из экспоненциального семейства

$$p(\mathbf{x} \mid \eta) = h(\mathbf{x})g(\eta)e^{\eta^\top \mathbf{u}(\mathbf{x})}$$

можно сразу оптом найти сопряжённые априорные распределения:

$$p(\eta \mid \chi, \nu) = f(\chi, \nu)g(\eta)^\nu e^{\nu\eta^\top \chi},$$

где χ – гиперпараметры, а g то же самое, что в исходном распределении.

Упражнение. Проверьте это и получите вышеописанные примеры как частные случаи.

Когда и μ, и σ^2 меняются

- В настоящем сопряжённом априорном распределении будут:

$$\begin{aligned} x \mid \mu, \tau &\sim \mathcal{N}(\mu, \tau), \\ \mu \mid \tau &\sim \mathcal{N}(\mu_0, n_0\tau), \\ \tau &\sim G(\alpha, \beta). \end{aligned}$$

- Давайте выясним, как изменятся параметры, и заодно докажем.

Когда и μ, и σ^2 меняются

- Самое простое – это, по уже известным результатам,

$$\mu \mid x, \tau \sim \mathcal{N}\left(\frac{n\tau}{n\tau + n_0\tau}\bar{x} + \frac{n_0\tau}{n\tau + n_0\tau}\mu_0, n\tau + n_0\tau\right).$$

- Затем давайте разберёмся с $\tau \mid x$:

$$p(\tau, \mu \mid x) \propto p(\tau) \cdot p(\mu \mid \tau) \cdot p(x \mid \tau, \mu),$$

  и мы хотим это распределение маргинализовать по μ...

Когда и μ, и σ^2 меняются

- Подсчитаем:

$$
\begin{aligned}
p(\tau, \mu \mid x) &\propto p(\tau) \cdot p(\mu \mid \tau) \cdot p(x \mid \tau, \mu) \\
&\propto \tau^{\alpha-1} e^{-\tau\beta} \cdot \tau^{\frac{1}{2}} e^{-\frac{n_0 \tau}{2}(\mu-\mu_0)^2} \cdot \tau^{\frac{n}{2}} e^{-\frac{\tau}{2}\sum(x_i-\mu)^2} \\
&\propto \tau^{\alpha+\frac{n}{2}-\frac{1}{2}} e^{-\tau\left(\beta+\frac{1}{2}\sum(x_i-\bar{x})^2\right)} e^{-\frac{\tau}{2}\left(n_0(\mu-\mu_0)^2+n(\bar{x}-\mu)^2\right)}
\end{aligned}
$$

(простой трюк: $x_i - \mu = x_i - \bar{x} + \bar{x} - \mu$).

Когда и μ, и σ^2 меняются

- Теперь надо проинтегрировать

$$\int_{\mu} e^{-\frac{\tau}{2}\left(n_0(\mu-\mu_0)^2+n(\bar{x}-\mu)^2\right)} d\mu.$$

Упражнение. Проинтегрируйте. :) Должна получиться нормировочная константа

$$\tau^{-\frac{1}{2}} e^{\frac{-nn_0\tau}{2(n+n_0)}(\bar{x}-\mu_0)^2}.$$

Когда и μ, и σ^2 меняются

- Таким образом, получается апостериорное распределение

$$p(\tau \mid x) \propto \tau^{\alpha+\frac{n}{2}-1} e^{-\tau\left(\beta+\frac{1}{2}\sum(x_i-\bar{x})^2+\frac{nn_0}{2(n+n_0)}(\bar{x}-\mu_0)^2\right)}.$$

- Итого результаты такие:

$$\begin{aligned} \mu \mid \tau, x &\sim \mathcal{N}\left(\tfrac{n\tau}{n\tau+n_0\tau}\bar{x}+\tfrac{n_0\tau}{n\tau+n_0\tau}\mu_0, n\tau+n_0\tau\right), \\ \tau \mid x &\sim G\left(\alpha+\tfrac{n}{2}, \beta+\tfrac{1}{2}\sum(x_i-\bar{x})^2+\tfrac{nn_0}{2(n+n_0)}(\bar{x}-\mu_0)^2\right). \end{aligned}$$

Предсказание

- Теперь предсказание нового x_{new}:

$$p(x_{\text{new}} \mid x) = \int\int \underbrace{\text{Gamma}}_{\tau|x} \cdot \underbrace{\text{Gaussian}}_{\mu|\tau,x} \cdot \underbrace{\text{Gaussian}}_{x_{\text{new}}|\tau,\mu} d\tau d\mu =$$

$$= \int \underbrace{\text{Gamma}}_{\tau|x} \int \underbrace{\text{Gaussian}}_{\mu|\tau,x} \cdot \underbrace{\text{Gaussian}}_{x_{\text{new}}|\tau,\mu} d\tau d\mu =$$

$$= \int \underbrace{\text{Gamma}}_{\tau|x} \cdot \underbrace{\text{Gaussian}}_{x_{\text{new}}|\tau,x} d\tau = \ldots$$

- В результате получится распределение Стьюдента.

Спасибо!

Спасибо за внимание!