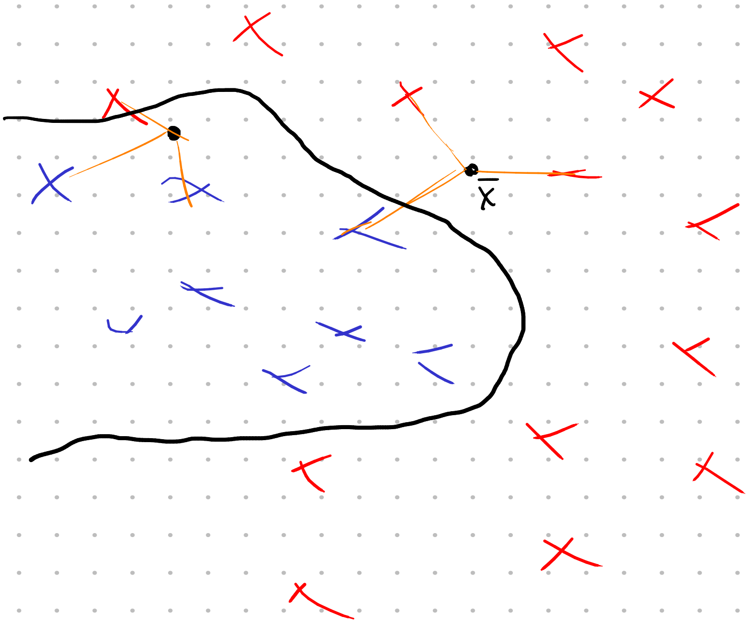
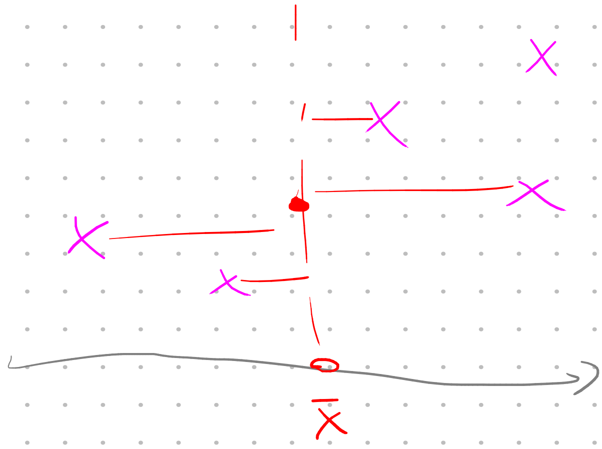
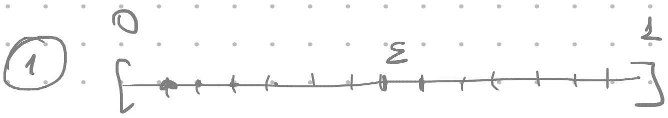
# Nearest neighbors

$k$

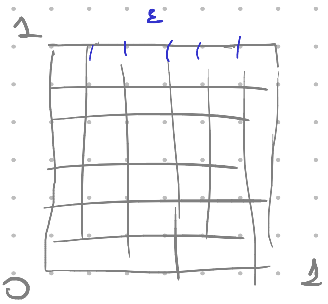



## Curse of dimensionality



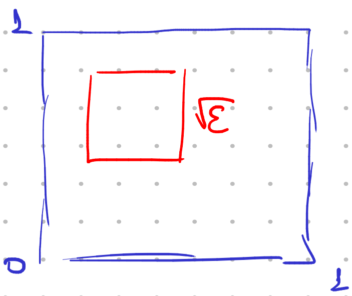

$$\int_0^1 f(x)\,dx \approx \sum^{1/\varepsilon}$$

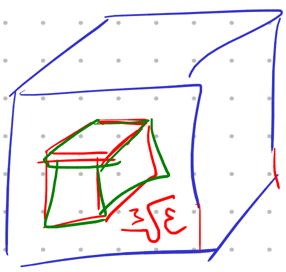


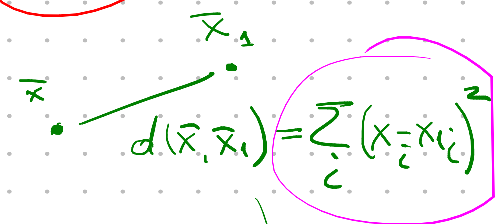

$1/\varepsilon^2$

$1/\varepsilon^d$


$\varepsilon^{1/d}$

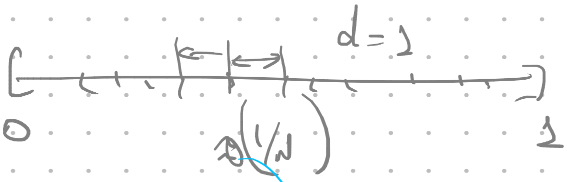



$$d(\bar{x}, x_1) = \sum_i (\bar{x}_i - x_{1i})^2$$

$d=100$




$O(1/N)$

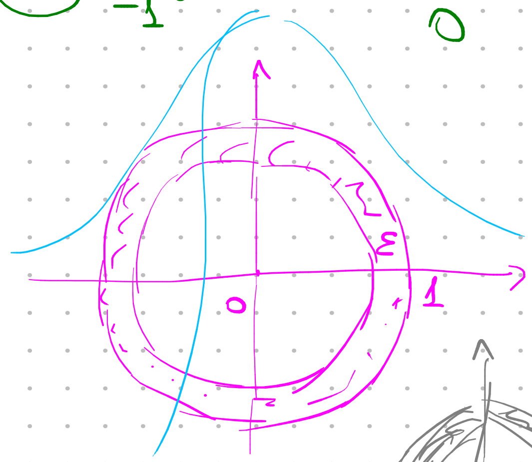
N torek

$$O\left(N^{-1/d}\right)$$

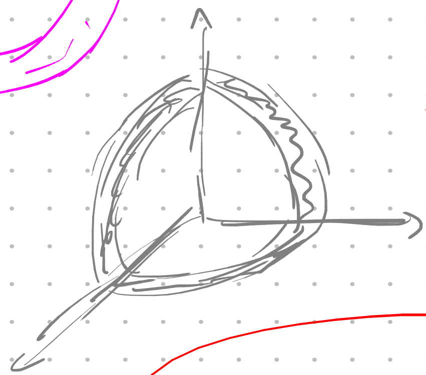
$d = 100$ N torek

3


ε



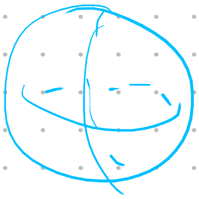

$$\frac{\pi\cdot 1^2 - \pi(1-\varepsilon)^2}{\pi\cdot 1^2} = 1-(1-\varepsilon)^2 = 2\varepsilon-\varepsilon^2$$



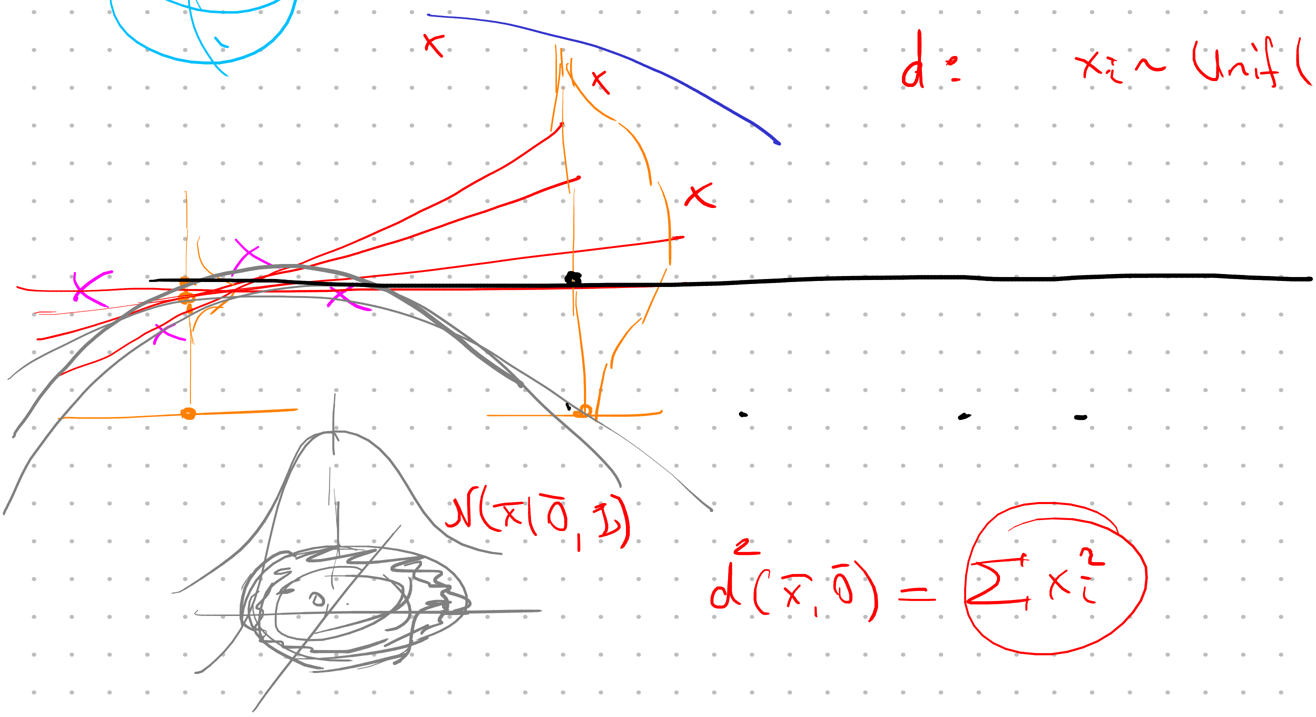
$$\frac{\cancel{\frac{4}{3}\pi} r^3 - \cancel{\frac{4}{3}}\pi(1-\varepsilon)^3 \cancel{r^3}}{\cancel{\frac{4}{3}\pi r^3}} =$$

$$= 1-(1-\varepsilon)^3$$

$$1-(1-\varepsilon)^d \xrightarrow[d\to\infty]{} 1$$



$d$: $\quad x_i \sim \mathrm{Unif}([-1,1])$




$$d^2(\bar{x},\bar{0}) = \sum_i x_i^2$$

## Statistical decision theory

$\bar{x}\in\mathbb{R}^d$, $y\in\mathbb{R}$ $\qquad D=\{(\bar{x}_n,y_n)\}_{n=1}^N$

$p(y|\bar{x})$ $\qquad p(\bar{x},y) = p(\bar{x})\,p(y|\bar{x})$

$$L(y,f(\bar{x})) = (y-f(\bar{x}))^2$$

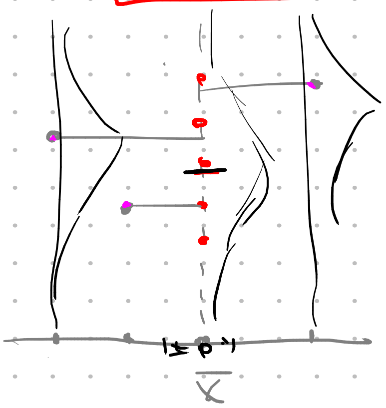
$f \nearrow \min$

$$EPE[f] = E_{p(\bar x, y)}[L(y, f(\bar x))] = \iint (y - f(\bar x))^2\, p(\bar x, y)\, d\bar x\, dy =$$

$$= \iint (y - f(\bar x))^2\, p(y|\bar x)\, p(\bar x)\, dy\, d\bar x = \int \Big[ \underbrace{\int (y - f(\bar x))^2\, p(y|\bar x)\, dy}_{\xrightarrow[f(\bar x)]{} \min} \Big] p(\bar x)\, d\bar x$$

$$\boxed{\hat f(\bar x) = E_{p(y|\bar x)}[y]}$$ – regression function

$$\int (x-a)^2\, p(x)\, dx = \int x^2 p(x)\, dx - 2a\,Ex + a^2$$



$$\hat f = E_{p(y|\bar x)}[y]$$

$$\approx \frac{1}{N} \sum_n y_n, \quad y_n \sim p(y|\bar x)$$

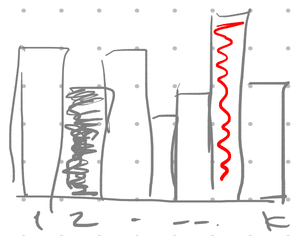
$$\approx \frac{1}{N} \sum_n y_n, \quad y_n \sim N\text{-}NN(\bar x)$$

$C_1 \dots C_K$ $\quad C_1 - C_K$

$$L(y, f(\bar x)) = [y \neq f(\bar x)] = \begin{cases} 0, & y = f(\bar x) \\ 1, & y \neq f(\bar x) \end{cases}$$

$$EPE[f] = \iint L(y, f(\bar x))\, p(\bar x, y)\, dy\, d\bar x =$$

$$= \int \Big[ \underbrace{\sum_{k=1}^{K} L(C_k, f(\bar x))\, p(C_k|\bar x)}_{\xrightarrow[f(\bar x)]{} \min} \Big] p(\bar x)\, d\bar x$$




$p(C_k|\bar x)$

$$\boxed{\hat f(\bar x) = \arg\max_k p(C_k|\bar x)}$$ ← optimal Bayes classifier

$L(y, f(\bar x))$

| Dis \ Test | pos | neg |
|---|---|---|
| pos | 0 | 1000 |
| neg | 1 | 0 |

$$\boxed{\hat f(\bar x) = \arg\max_s \sum_{k=1}^{K} L(C_k, s)\, p(C_k|\bar x)}$$

$$EPE[f] = \iint (y - f(\bar{x}))^2 p(\bar{x}, y)\, d\bar{x}\, dy =$$

$$\pm \hat{f}(\bar{x}) = \mathbb{E}_{p(y|\bar{x})}[y]$$

$$= \iint \left( (y - \hat{f}(\bar{x}))^2 - 2(y - \hat{f}(\bar{x}))(\hat{f}(\bar{x}) - f(\bar{x})) + (\hat{f}(\bar{x}) - f(\bar{x}))^2 \right) p(\bar{x}, y)\, d\bar{x}\, dy$$

$$= \mathbb{E}_{p(\bar{x},y)}[(y - \hat{f}(\bar{x}))^2] - \cancel{2\,\underbrace{\mathbb{E}_{p(\bar{x},y)}[(y - \hat{f})(\hat{f} - f)]}} + \mathbb{E}_{p(\bar{x},y)}[(\hat{f}(\bar{x}) - f(\bar{x}))^2]$$

$$p(\bar{x},y) = p(\bar{x})\,p(y|\bar{x})$$

$$= \mathbb{E}_{p(\bar{x})}[(\hat{f}(\bar{x}) - f(\bar{x})) \cdot \overbrace{\mathbb{E}_{p(y|\bar{x})}[y - \hat{f}(\bar{x})]}^{=0}]$$

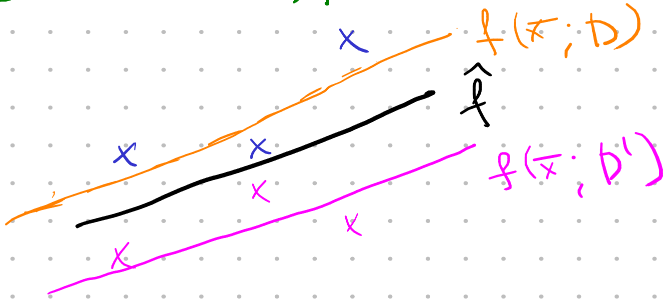
$$EPE[f] = \overbrace{\mathbb{E}[(y - \hat{f}(\bar{x}))^2]}^{\text{Noise}} + \underline{\mathbb{E}[(\hat{f}(\bar{x}) - f(\bar{x}))^2]}$$

$$\pm \mathbb{E}_D f(\bar{x})$$

$$\iint (\hat{f} - f)^2 p(\bar{x}, y)\, d\bar{x}\, dy = \int (\hat{f}(\bar{x}) - f(\bar{x}))^2 p(\bar{x})\, d\bar{x} = (*)$$

$$\boxed{f(\bar{x}; D)}, \quad D \sim p(\bar{x}, y) = \{(\bar{x}_n, y_n)\}_{n=1}^{N}$$

$$\mathbb{E}_D f(\bar{x})$$




$$(*) = \int (\hat{f}(\bar{x}) - \mathbb{E}_D f(\bar{x}))^2 p(\bar{x})\, d\bar{x} - \cancel{2\int (\hat{f} - \mathbb{E}_D f)(\mathbb{E}_D f - f) p(\bar{x})\, d\bar{x}} + \int (\mathbb{E}_D f - f)^2 p(\bar{x})\, d\bar{x}$$

$$EPE[f] = \mathbb{E}[(\hat{f}(\bar{x}) - \mathbb{E}_D f(\bar{x}))^2] \quad \text{Bias}$$

$$+ \mathbb{E}[(\mathbb{E}_D f(\bar{x}) - f(\bar{x}; D))^2] \quad \text{Variance}$$

$$+ \mathbb{E}[(y - \hat{f}(\bar{x}))^2] \quad \text{Noise}$$

# Байєсовий вивод для гаусіана

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$\Lambda = \Sigma^{-1}$$

$\tau = \frac{1}{\sigma^2}$ – precision

$$\mathcal{N}(x|\mu,\tau) = \sqrt{\frac{\tau}{2\pi}}\, e^{-\frac{\tau}{2}(x-\mu)^2}$$

$$\mathcal{N}(x_1,\dots,x_n|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

① $\tau = \text{Const}$ $\quad \mu$

$$\log p(x_1 - x_n|\mu) = \text{Const} - \frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$p(\mu) = \mathcal{N}(\mu|\mu_0,\tau_0) = \sqrt{\frac{\tau_0}{2\pi}}\, e^{-\frac{\tau_0}{2}(\mu-\mu_0)^2}$$

$$x_i^2 - 2x_i\mu + \mu^2$$

$$\log p(\mu|x_1 \dots x_n) = \text{Const} - \frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{\tau_0}{2}(\mu-\mu_0)^2 =$$

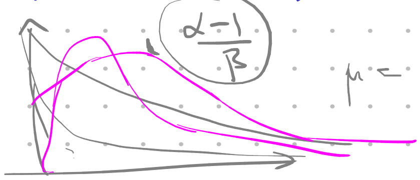
$$= \text{Const} - \frac{1}{2}\mu^2(\tau_0 + \tau\cdot n) + \mu\left(\mu_0\tau_0 + \tau\cdot\sum_{i=1}^{n}x_i\right) - \frac{\tau}{2}\sum_i x_i^2$$

$$= \text{Const} - \frac{\tau_0+\tau n}{2}\left(\mu - \frac{\tau_0\mu_0 + \tau\cdot\sum x_i}{\tau_0 + n\tau}\right)^2$$

$$\frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$$

$$p(\mu|x_1 - x_n) = \mathcal{N}\left(\mu \,\middle|\, \frac{\tau_0\mu_0 + \tau\sum x_i}{\tau_0 + n\tau},\ \tau_0 + n\tau\right)$$

$$\sigma_n^2 = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}$$




$\mu = \alpha/\beta \quad \sigma^2 = \alpha/\beta^2$

② $\mu = \text{const}, \quad \tau$

$$\log p(x_1 - x_n|\tau) = \text{Const} + \frac{n}{2}\log\tau - \frac{\tau}{2}\sum_i (x_i-\mu)^2$$

$$p(\tau|\alpha_0,\beta_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\tau^{\alpha_0-1}e^{-\beta_0\tau} \qquad \tau \in (0,+\infty)$$

– gamma distribution

$$\log p(\tau \mid x_1 \dots x_n) = \text{Const} + \frac{n}{2}\log\tau - \frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$+ (\alpha_0 - 1)\log\tau - \beta_0\tau$$

$$p(\tau \mid x_1 \dots x_n) = \text{Gamma}\left(\tau \,\middle|\, \alpha_0 + \frac{n}{2}, \beta_0 + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

3 $\mu, \tau = ?$ $\quad \log p(x_1 \dots x_n \mid \mu, \tau) = \text{Const} + \frac{n}{2}\log\tau - \frac{\tau}{2}\sum_i (x_i - \mu)^2$

$$\cancel{p(\mu,\tau) = p(\mu)\,p(\tau) = \mathcal{N}(\mu \mid \mu_0, \tau_0)\cdot \text{Gam}(\tau \mid \alpha_0, \beta_0)}$$

$$\cancel{\log p(\mu,\tau \mid x_1 \dots x_n) = \text{Const} + \frac{n}{2}\log\tau - \frac{\tau}{2}\sum(x_i-\mu)^2}$$

$$\cancel{-\frac{\tau_0}{2}(\mu-\mu_0)^2 + (\alpha_0-1)\log\tau - \beta_0\tau =}$$

$$\cancel{= \text{Const} + \left(\alpha_0 + \frac{n}{2} - 1\right)\log\tau - (\_\_)\tau - (\_ \_\mu \_ \_\mu^2)}$$

$$p(\mu,\tau) = \underbrace{p(\tau)}_{\text{Gam}}\,\underbrace{p(\mu \mid \tau)}_{\mathcal{N}(\mu \mid \mu_0, \lambda\tau)} = \underbrace{p(\mu)}_{\mathcal{N}}\,\underbrace{p(\tau \mid \mu)}_{\text{Gam}(\mu?)}$$

$$p(\mu,\tau) = \text{Gam}(\tau \mid \alpha_0, \beta_0)\cdot\mathcal{N}(\mu \mid \mu_0, \lambda_0\tau)$$

$$\log p(\mu,\tau \mid x_1 \dots x_n) = \text{Const} + \log p(x_1 \dots x_n \mid \mu,\tau) + \log p(\mu,\tau) =$$

$$= \text{Const} + \frac{n}{2}\log\tau - \frac{\tau}{2}\sum_i (x_i-\mu)^2 + (\alpha_0-1)\log\tau - \beta_0\tau$$

$$+ \frac{1}{2}\log(\lambda_0\tau) - \frac{\lambda_0\tau}{2}(\mu-\mu_0)^2 =$$

$$= \text{Const} + \left(\alpha_0 + \frac{n}{2} - 1\right)\log\tau + \frac{1}{2}\log\tau - \beta_0\tau -$$

$$-\frac{\tau}{2}\left(\sum_i (x_i-\mu)^2 + \lambda_0(\mu-\mu_0)^2\right) = (*)$$

$$\sum_i x_i^2 - 2\mu\cdot\sum x_i + \mu^2\cdot n + \lambda_0\mu^2 - 2\lambda_0\mu_0\mu + \lambda_0\mu_0^2 =$$

$$= \mu^2(\lambda_0+n) - 2\mu\left(\lambda_0\mu_0 + \sum x_i\right) + \lambda_0\mu_0^2 + \sum_i x_i^2 =$$

$$= (\lambda_0+n)\left(\mu - \frac{\lambda_0\mu_0+\sum x_i}{\lambda_0+n}\right)^2 - \cancel{(\lambda_0+n)}\cdot\frac{(\lambda_0\mu_0+\sum x_i)^2}{(\lambda_0+n)^{\cancel{2}}} + \lambda_0\mu_0^2+\sum x_i^2$$

$$(*) = \text{Const} + \overbrace{\left(\alpha_0 + \frac{n}{2} - 1\right)\log\tau} + \frac{1}{2}\log\tau - \beta_0\tau - \frac{(\lambda_0+n)\tau}{2}\left(\mu - \frac{\lambda_0\mu_0+\sum x_i}{\lambda_0+n}\right)^2 - \frac{\tau}{2}\left(\lambda_0\mu_0^2 + \sum x_i^2 - \frac{(\lambda_0\mu_0+\sum x_i)^2}{\lambda_0+n}\right)$$

$$\text{Const} + \log\sqrt{\frac{(\lambda_0+n)\tau}{2\pi}}\cdot e^{-\frac{(\lambda_0+n)\tau}{2}(\_\_)^2}$$

$$p(\mu,\tau|x_1-x_n) = \text{Gam}\left(\tau \,\middle|\, \overbrace{\alpha_0+\frac{n}{2}}^{\alpha_n},\ \overbrace{\beta_0 + \frac{\tau}{2}\left(\lambda_0\mu_0^2+\sum x_i^2 - \frac{(\lambda_0\mu_0+\sum x_i)^2}{\lambda_0+n}\right)}^{\beta_n}\right)$$

$$\times\ \mathcal{N}\left(\mu \,\middle|\, \overbrace{\frac{\lambda_0\mu_0+\sum_i x_i}{\lambda_0+n}}^{\mu_n},\ \overbrace{(\lambda_0+n)}^{\lambda_n}\tau\right)$$

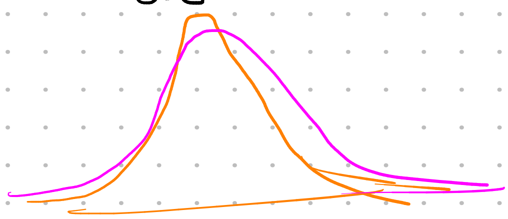
(4) $$p(\mu|x_1-x_n) = \int_0^\infty p(\mu,\tau|x_1-x_n)\,d\tau =$$

$$= \int_0^\infty \text{Gam}(\tau|\alpha_n,\beta_n)\cdot\mathcal{N}(\mu|\mu_n,\lambda_n\tau)\,d\tau =$$

$$= \text{Const}\cdot\int_0^\infty \tau^{\alpha_n-1}e^{-\beta_n\tau}\cdot\tau^{\frac{1}{2}}\cdot e^{-\frac{\lambda_n\tau}{2}(\mu-\mu_n)^2}\,d\tau =$$

$$= \text{Const}\cdot\int_0^\infty \tau^{(\alpha_n+\frac{1}{2})-1}\cdot e^{-\tau\cdot\left(\beta_n+\frac{\lambda_n}{2}(\mu-\mu_n)^2\right)}\,d\tau =$$

$$= \text{Const}\cdot\frac{\Gamma(\alpha_n+\frac{1}{2})}{\left(\beta_n+\frac{\lambda_n}{2}(\mu-\mu_n)^2\right)^{\alpha_n+\frac{1}{2}}}$$



Student's t-distribution

$$p(x) = \text{Const} \cdot \left(1 + \frac{x^2}{D}\right)^{-\frac{D+1}{2}}$$

$$p(\mu | x_1 - x_n) = \text{Const} \cdot \left(1 + \frac{\lambda_n}{2\beta_n}(\mu - \mu_n)^2\right)^{-\frac{2\alpha_n+1}{2}}$$

(5) **Predictive distribution**

$$p(x | x_1 - x_n) = \int_0^\infty \int_{-\infty}^{\infty} \mathcal{N}(x | \mu, \tau) \cdot \text{Gam}(\tau | \alpha_n, \beta_n) \cdot \mathcal{N}(\mu | \mu_n, \lambda_n \tau) \, d\mu \, d\tau$$

?

$$\int \mathcal{N}(x | \mu, \tau) \, \mathcal{N}(\mu | \mu_n, \lambda_n \tau) \, d\mu = \mathcal{N}(x | \mu', \tau')$$

$$C + \frac{1}{2}\log\tau + \frac{1}{2}\log\tau - \frac{\tau}{2}(x-\mu)^2 - \frac{\lambda_n\tau}{2}(\mu - \mu_n)^2 =$$

$$\text{Const} + \log\tau - \frac{\tau}{2}x^2 - \frac{\lambda_n\tau}{2}\mu_n^2 - \frac{\mu^2}{2}(\tau + \lambda_n\tau) + \mu\tau(x + \lambda_n\mu_n) =$$

$$= \underbrace{\text{Const} + \frac{1}{2}\log\tau - \frac{\tau}{2}x^2 - \frac{\lambda_n\tau}{2}\mu_n^2 + \frac{\tau}{2}\frac{(x + \lambda_n\mu_n)^2}{\lambda_n + 1}}_{\frac{\tau}{2(\lambda_n+1)}(x^2 + 2x\lambda_n\mu_n + \lambda_n^2\mu_n^2)} \cancel{- \frac{\tau}{2}\left(\mu - \frac{x + \lambda_n\mu_n}{\lambda_n + 1}\right)^2 + \frac{1}{2}\log\tau}$$

$$= \text{Const} + \frac{1}{2}\log\tau - \frac{\tau}{2} \cdot \frac{\lambda_n}{\lambda_n + 1} \cdot x^2 + \tau \cdot \frac{\lambda_n}{\lambda_n + 1}\mu_n \cdot x + \frac{\tau\lambda_n^2\mu_n^2}{2(\lambda_n + 1)} - \frac{\lambda_n\tau\mu_n^2}{2}$$

$$= \underbrace{\text{Const} + \frac{1}{2}\log\tau - \frac{\tau}{2} \cdot \frac{\lambda_n}{\lambda_n + 1}(x - \mu_n)^2}_{} + \cancel{\frac{\tau}{2}\frac{\mu_n^2\lambda_n}{\lambda_n + 1} + \frac{\tau}{2}\frac{\lambda_n^2\mu_n^2}{\lambda_n + 1} - \frac{\tau}{2}\lambda_n\mu_n^2}$$

$$\mathcal{N}\left(x \,\middle|\, \mu' = \mu_n, \ \tau' = \tau \cdot \frac{\lambda_n}{\lambda_n + 1}\right)$$

$$p(x \mid x_1 \ldots x_n) = \int_0^\infty \mathcal{N}\left(x \mid \mu_n, \tau \cdot \frac{\lambda_n}{\lambda_n+1}\right) \cdot \mathrm{Gam}(\tau \mid \alpha_n, \beta_n) \, d\tau$$

$$= \mathrm{Const} \cdot \frac{\Gamma\left(\alpha_n + \frac{1}{2}\right)}{\left(\beta_n + \frac{\lambda_n}{2(\lambda_n+1)} \cdot (\mu - \mu_n)^2\right)^{\alpha_n + \frac{1}{2}}}$$