№1.

$x_1, \dots, x_n \sim \mathcal{N}(x | \mu_0, \tau^{-1})$

$p(\tau | k_0, \theta_0) = \frac{1}{\Gamma(k_0)\theta_0^{k_0}} \tau^{k_0 - 1} \exp\left(-\frac{\tau}{\theta_0}\right)$

$p(\tau | X) = ?$

Решение:

По теореме Байеса

$$p(\tau / X) = \frac{p(X|\tau)\, p(\tau)}{p(X)} = C p(X|\tau) p(\tau)$$

$$p(\tau / X) \propto p(X|\tau) p(\theta)$$

$$p(X|\tau) = \prod_{i=1}^{n} p(x_i | \mu_0, \tau^{-1}) = \frac{1}{(2\pi \tau^{-1})^{n/2}} e^{-\frac{1}{2\tau^{-1}} \sum (x_i - \mu_0)^2}$$

$$p(\tau / X) \propto \frac{1}{(2\pi \tau^{-1})^{n/2}} e^{-\frac{1}{2\tau^{-1}} \sum (x_i - \mu_0)^2} \quad p(\theta) =$$

$$= \underbrace{\frac{1}{(2\pi \tau^{-1})^{n/2}}}_{const} e^{-\frac{1}{2\tau^{-1}} \sum (x_i - \mu_0)^2 - \frac{1}{\theta_0 \tau^{-1}}} \cdot \underbrace{\frac{1}{\Gamma(k_0)\theta_0^{k_0}}}_{const} \tau^{k_0 - 1} \propto$$

$$\propto \tau^{(k_0 + \frac{n}{2}) - 1} e^{-\frac{1}{2}\left(\sum (x_i - \mu_0)^2 + \frac{1}{\theta_0}\right)\tau} I\{\tau > 0\} \propto$$

$$\propto \Gamma\left(\tau \,\middle|\, k_0 + \frac{n}{2}, -\frac{1}{2}\left(\sum (x_i - \mu_0)^2 + \frac{1}{\theta_0}\right)\right) = p(\tau / X)$$

№2.

$$p(y \mid a(x)) = \frac{(a(x))^y e^{-a(x)}}{y!}$$

$$\ln p(y \mid a(x)) = y \ln[a(x)] - a(x) - \underbrace{\ln y!}_{const} \to max$$

Ответ: $a(x) - y \ln[a(x)] \to min$

№3.

$$p(Y|X,w) = \mathcal{N}(Y|Xw, \beta I)$$

$$p(w) = \mathcal{N}(w|0, \alpha I)$$

$$p(w|X,Y) = ?$$

$$p(y_*|x_*, X, Y) = ?$$

$$-\log p(y|X,w) - \log p(w) = -\sum_{i=1}^{N} p(y_i|x_i,w) - \sum_{j=1}^{D} p(w_j) =$$

$$= -\sum_{i=1}^{N} \left(-\frac{1}{2}\ln(2\pi\beta) - \frac{(y_i - (w,x_i))^2}{2\beta}\right) -$$

$$- \sum_{j=1}^{D} \left(-\frac{1}{2}\log(2\pi\alpha) - \frac{w_j^2}{2\alpha}\right) =$$

$$= \|Y - Xw\|^2 + \frac{\beta}{\alpha}\|w\|^2 + const \to \min$$

Линейная регрессия с L_2-регуляризацией

$$\hat{w} = \left(X^T X + \frac{\beta}{\alpha} I\right)^{-1} X^T y$$

$$\log p(w|X,Y) = \frac{1}{2\beta}\|Y - Xw\|^2 + \frac{1}{2\alpha}\|w\|^2 + const =$$

$$= \frac{1}{2\beta}(y^T y - w^T X^T y - y^T wx + w^T X^T Xw) + \frac{1}{2\alpha} w^T w + \tilde{C} =$$

$$= w^T\left(\frac{1}{2\beta} X^T X + \frac{1}{2\alpha} I\right) w - \frac{1}{2\beta} w^T X^T y - \frac{1}{2\beta} y^T wx + \tilde{C} =$$

$$= \frac{1}{2}(w - \hat{w})^T \left(\frac{1}{\beta} x^T x + \frac{1}{\alpha} I\right)(w - \hat{w}) + \tilde{C}$$

$$p(w \mid X, y) = \mathcal{N}\left(\left(X^T X + \frac{\beta}{\alpha}\right)^{-1} X^T y, \left(\frac{1}{\beta} X^T X + \frac{1}{\alpha} I\right)^{-1}\right)$$

$$p(y_* \mid x_*, X, Y) = \int p(y_* \mid x_*, w)\, p(w \mid X, Y)\, dw =$$

$$= \mathcal{N}\left(y_* \mid x_* \hat{w}, \beta + \beta x_*^T \left(X^T X + \frac{\beta}{\alpha} I\right)^{-1} x_*\right)$$