Курузов Илья, 678

Задание 3

1.

$$\frac{\partial J(\mathbf{U},\mathbf{V})}{\partial U_{ij}} = \frac{\partial}{\partial U_{ij}} \sum_{l=1}^{n} \left(\sum_{k=1}^{n} U_{ik} V_{kl} - Y_{il} \right)^2 + \lambda U_{ij}$$

$$\frac{\partial J(\mathbf{U},\mathbf{V})}{\partial U_{ij}} = 2 \sum_{l=1}^{n} V_{jl} \left(\sum_{k=1}^{n} U_{ik} V_{kl} - Y_{il} \right) + \lambda U_{ij}$$

Свернем в матричный вид:

$$\boxed{\frac{\partial J(\mathbf{U},\mathbf{V})}{\partial \mathbf{U}} = 2\left(\mathbf{U}\mathbf{V}\mathbf{V}^\top - \mathbf{Y}\mathbf{V}\right) + \lambda \mathbf{U}}$$

Функция симметричная по $\mathbf{U}$ и $\mathbf{V}$, их этих соображений, получаем выражение для производной по другой матрице:

$$\boxed{\frac{\partial J(\mathbf{U},\mathbf{V})}{\partial \mathbf{V}} = 2\left(\mathbf{V}\mathbf{U}\mathbf{U}^\top - \mathbf{Y}\mathbf{U}\right) + \lambda \mathbf{V}}$$

2.

$$\frac{\partial f(\mathbf{w})}{\partial w_k} = \sum_{i=1}^{m} \frac{\partial}{\partial w_k} \log\left(1 + e^{-y_i \mathbf{w}^\top \mathbf{x}_i}\right)$$

$$\frac{\partial f(\mathbf{w})}{\partial w_k} = \sum_{i=1}^{m} \frac{-y_i x_i^k e^{-y_i \mathbf{w}^\top \mathbf{x}}}{1 + e^{-y_i \mathbf{w}^\top \mathbf{x}_i}}$$

$$\boxed{\operatorname{grad} f(\mathbf{w}) = -\sum_{i=1}^{m} \frac{-y_i \mathbf{x} e^{y_i \mathbf{w}^\top \mathbf{x}}}{1 + e^{-y_i \mathbf{w}^\top \mathbf{x}_i}}}$$

Пришло время для гессиана:

$$H_{mk}(f) = \frac{\partial}{\partial w_m} \frac{\partial f(\mathbf{w})}{\partial w_k} = \frac{\partial}{\partial w_m} \left(\sum_{i=1}^{m} \frac{-y_i x_i^k e^{-y_i \mathbf{w}^\top \mathbf{x}}}{1 + e^{-y_i \mathbf{w}^\top \mathbf{x}_i}} \right)$$

$$H_{mk}(f) = -\sum_{i=1}^{m} y_i x_i^k \frac{\partial}{\partial w_m} \frac{e^{-y_i \mathbf{w}^\top \mathbf{x}}}{1 + e^{-y_i \mathbf{w}^\top \mathbf{x}_i}}$$

$$H_{mk}(f) = -\sum_{i=1}^{m} y_i x_i^k \frac{-y_i x_i^m e^{-y_i \mathbf{w}^\top \mathbf{x}_i}(1 + e^{-y_i \mathbf{w}^\top \mathbf{x}_i}) + y_i x_i^m e^{-2y_i \mathbf{w}^\top \mathbf{x}_i}}{(1 + e^{-y_i \mathbf{w}^\top \mathbf{x}_i})^2}$$

$$H_{mk}(f) = \sum_{i=1}^{m} \frac{y_i^2 x_i^k x_i^m e^{-y_i \mathbf{w}^\top \mathbf{x}_i}}{(1 + e^{-y_i \mathbf{w}^\top \mathbf{x}_i})^2}$$

Возвращаясь к матрицам и векторам:

$$\boxed{H_{mk}(f) = \sum_{i=1}^{m} \frac{y_i^2 e^{-y_i \mathbf{w}^\top \mathbf{x}_i}}{(1 + e^{-y_i \mathbf{w}^\top \mathbf{x}_i})^2} \mathbf{x}_i \mathbf{x}_i^\top}$$

3.

$$J_{km} = \frac{\partial f_k}{\partial x_m} = \frac{\delta_{km} e^{w_k} \left(\sum\limits_{j=1}^{n} e^{w_j} \right) - e^{w_k + w_m}}{\left(\sum\limits_{j=1}^{n} e^{w_j} \right)^2}$$

Я не могу увидеть, как это красиво переписать в матричном виде, а поэтому пусть так и остается:

$$\boxed{J_{km} = \frac{\partial f_k}{\partial x_m} = \frac{\delta_{km} e^{w_k} \left(\sum\limits_{j=1}^{n} e^{w_j} \right) - e^{w_k + w_m}}{\left(\sum\limits_{j=1}^{n} e^{w_j} \right)^2}}$$

4.

$$\frac{\partial R(\mathbf{x}|\mathbf{A})}{\partial \mathbf{x}} = \frac{(\mathbf{A} + \mathbf{A}^\top)\mathbf{x} \cdot \mathbf{x}^\top\mathbf{x} - \mathbf{x}^\top\mathbf{A}\mathbf{x} \cdot 2\mathbf{x}}{(\mathbf{x}^\top\mathbf{x})^2} =$$

$$= \frac{2\mathbf{A}\mathbf{x} \cdot \mathbf{x}^\top\mathbf{x} - \mathbf{x}^\top\mathbf{A}\mathbf{x} \cdot 2\mathbf{x}}{(\mathbf{x}^\top\mathbf{x})^2} = 2\frac{\mathbf{A}\mathbf{x} \cdot \mathbf{x}^\top\mathbf{x} - \mathbf{x}^\top\mathbf{A}\mathbf{x} \cdot \mathbf{x}}{(\mathbf{x}^\top\mathbf{x})^2}$$

$$\boxed{\operatorname{grad} R(\mathbf{x}|\mathbf{A}) = 2\frac{\mathbf{A}\mathbf{x} \cdot \mathbf{x}^\top\mathbf{x} - \mathbf{x}^\top\mathbf{A}\mathbf{x} \cdot \mathbf{x}}{(\mathbf{x}^\top\mathbf{x})^2}}$$

При x совпадающим с собственным вектором градиент обнуляется. Причем максимум, достигается на максимальном собственном числе (видно из вида функции отношения Рэлея).

5.

1) Из курса линейной алгебры, известно, что

$$f(\mathbf{X}) = \sum_{i=1}^{n} \lambda_i = \operatorname{tr} \mathbf{X} = \sum_{i=1}^{n} x_{ii}$$

$$\boxed{\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{E}}$$

2) Из курса линейной алгебры, известно, что

$$f(\mathbf{X}) = \prod_{i=1}^{n} \lambda_i = \det \mathbf{X}$$

$$\boxed{\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{X}^{-\top} \det \mathbf{X}}$$