

Теорема 1. Пусть \mathcal{L} — дифференцируемая функция, такая что все стационарные точки \mathcal{L} являются локальными минимумами. Пусть также гессиан \mathbf{H}^{-1} функции потерь \mathcal{L} является обратимым в каждой стационарной точке, тогда:

$$\nabla_{\mathbf{h}} \mathcal{Q}(\mathbf{T}(\boldsymbol{\Theta}_0, \mathbf{h}), \mathbf{h}) = \nabla_{\mathbf{h}} \nabla_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\Theta}^\eta, \mathbf{h})^\top \mathbf{H}^{-1} \nabla_{\boldsymbol{\Theta}} \mathcal{Q}(\boldsymbol{\Theta}^\eta, \mathbf{h}). \quad (1)$$

Доказательство.

$$\begin{aligned} \nabla_{\boldsymbol{\Theta}} \mathcal{L}(\mathbf{T}(\boldsymbol{\Theta}_0, \mathbf{h})) &= 0 \Rightarrow \\ \Rightarrow \nabla_{\mathbf{h}} (\nabla_{\boldsymbol{\Theta}} \mathcal{L}(\mathbf{T}(\boldsymbol{\Theta}_0, \mathbf{h}))) &= \nabla_{\boldsymbol{\Theta}, \mathbf{h}} \mathcal{L}(\boldsymbol{\Theta}^\eta, \mathbf{h}) + \nabla_{\boldsymbol{\Theta}}^2 \mathcal{L}(\boldsymbol{\Theta}^\eta, \mathbf{h}) \frac{\partial \boldsymbol{\Theta}}{\partial \mathbf{h}} = 0 \Rightarrow \\ \Rightarrow \frac{\partial \boldsymbol{\Theta}}{\partial \mathbf{h}} &= -(\nabla_{\boldsymbol{\Theta}}^2 \mathcal{L}(\boldsymbol{\Theta}^\eta, \mathbf{h}))^{-1} \nabla_{\boldsymbol{\Theta}, \mathbf{h}} \mathcal{L}(\boldsymbol{\Theta}^\eta, \mathbf{h}). \end{aligned} \quad (2)$$

$$\nabla_{\mathbf{h}} \mathcal{Q}(\mathbf{T}(\boldsymbol{\Theta}_0, \mathbf{h})) = \nabla_{\mathbf{h}} \mathcal{Q}(\boldsymbol{\Theta}^\eta, \mathbf{h}) + \nabla_{\boldsymbol{\Theta}} \mathcal{Q}(\boldsymbol{\Theta}^\eta, \mathbf{h})^\top \frac{\partial \boldsymbol{\Theta}}{\partial \mathbf{h}}. \quad (3)$$

Подставляя (2) в (3) получаем:

$$\begin{aligned} \nabla_{\mathbf{h}} \mathcal{Q}(\mathbf{T}(\boldsymbol{\Theta}_0, \mathbf{h})) &= \nabla_{\mathbf{h}} \mathcal{Q}(\boldsymbol{\Theta}^\eta, \mathbf{h}) - \nabla_{\boldsymbol{\Theta}} \mathcal{Q}(\boldsymbol{\Theta}^\eta, \mathbf{h})^\top (\nabla_{\boldsymbol{\Theta}}^2 \mathcal{L}(\boldsymbol{\Theta}^\eta, \mathbf{h}))^{-1} \nabla_{\boldsymbol{\Theta}, \mathbf{h}} \mathcal{L}(\boldsymbol{\Theta}^\eta, \mathbf{h}) = \\ &= \nabla_{\mathbf{h}} \mathcal{Q}(\boldsymbol{\Theta}^\eta, \mathbf{h}) - \nabla_{\boldsymbol{\Theta}, \mathbf{h}} \mathcal{L}(\boldsymbol{\Theta}^\eta, \mathbf{h})^\top (\nabla_{\boldsymbol{\Theta}}^2 \mathcal{L}(\boldsymbol{\Theta}^\eta, \mathbf{h}))^{-1} \nabla_{\boldsymbol{\Theta}} \mathcal{Q}(\boldsymbol{\Theta}^\eta, \mathbf{h}) \end{aligned} \quad (4)$$

□