3.4 №1

$x_i \in \mathbb{R}^d$, $Y_i \sim Bern(\mu_\theta(x_i)) \Leftrightarrow Bern(\sigma(x_i^T\theta))$

GD - ? IRLS - ?

Решение. В предположении, что эксперименты проводились независимо, имеем:

$$p_{\theta|y}(x|y) \propto p_{y|\theta}(y|x)\cdot p_\theta(x) = \prod_{i=1}^{n} p_{y_i|\theta}(y|x)\cdot p_\theta(x) \overset{y_i = x_i}{\underset{x_i=\theta}{=}} \prod_{i=1}^{n} [\sigma(x_i^T\theta)]^{y_i}[1-\sigma(x_i^T\theta)]^{1-y_i}\cdot \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det(\lambda^{-1}I_d)}}\cdot e^{-\frac{1}{2}\theta^T\lambda I_d\theta} \Rightarrow \arg\max_\theta p_{\theta|y}(x|y) \Leftrightarrow \arg\max_\theta \ln p_{\theta|y}(x|y)$$

$$\Rightarrow \ln p_{\theta|y}(x|y) = \sum_{i=1}^{n}[y_i \ln\sigma(x_i^T\theta) + (1-y_i)\ln(1-\sigma(x_i^T\theta))] + \left(-\frac{1}{2}\lambda\,\theta^T\theta\right) - \ln (2\pi)^{\frac{d}{2}}\sqrt{\det(\lambda^{-1}I_d)}$$

$$\frac{\partial \ln p_{\theta|y}(x|y)}{\partial\theta} = \sum_{i=1}^{n}[y_i - \sigma(x_i^T\theta)]x_i - \lambda\theta$$

Тогда шаг градиентного спуска:

$$\theta_{k+1} = \theta_k + \eta\cdot\left(\sum_{i=1}^{n}[y_i - \sigma(x_i^T\theta_k)]x_i - \lambda\theta_k\right)$$

Матр. вид:

$$\boxed{\theta_{k+1} = \theta_k + \eta\cdot\left(\underbrace{X^T}_{\begin{pmatrix} x_1^T \\ \vdots \\ x_n^T\end{pmatrix}}(y - \overbrace{s(\theta_k)}^{\sigma(x_i^T\theta_k)}) - \lambda\theta_k\right)}$$

Для IRLS имеем:

$$F(\theta) = -\ell(\theta) + \frac{1}{2}\lambda\theta^T\theta \Rightarrow \nabla F(\theta) = +X^T(s(\theta) - y) + \lambda\theta$$

$$\nabla\nabla F(\theta) = + X^T V(\theta) X + \lambda I_d, \text{ где } V(\theta) =$$

$$= diag\left(\sigma(x_i^T\theta)\cdot(1-\sigma(x_i^T\theta))\right)$$

Тогда получим формулу метода IRLS:

$$\theta_{k+1} = \theta_k - \left(X^T V(\theta_k) X + \lambda I_d\right)^{-1}\left(X^T(S(\theta_k) - y) + \lambda\theta_k\right)$$

Так как в ~~самой~~ задаче №2 просят в качестве функционала $+l(\theta)$ (при $\lambda=0$), то есть смысл переформулировать опт. задачу:

$$\boxed{\begin{array}{l} F(\theta) = +l(\theta) - \frac{1}{2}\lambda\theta^T\theta \longrightarrow \max\limits_{\theta} \\ \text{Тогда } \theta_{k+1} = \theta_k + \left(X^T V(\theta_k) X + \lambda I_d\right)^{-1}\left(-X^T(S(\theta_k) - y) - \lambda\theta_k\right) \end{array}}$$