VIR - Bonus Exercise - Batch Norm

Jiří Jirák

November 4, 2021

1 Task

Consider mini-batch $\mathbf{x} = [1\ 3]^{\mathrm{T}}$, affine function $y_i(\hat{x}_i) = \gamma\,\hat{x}_i + \beta$ where γ and β are the learnable parameters starting as $\gamma = 4$, $\beta = -2$. Consider Loss function defined as $\mathcal{L} = \sum_{i=1}^{n} y_i^2$.

1. Draw the computational graph of given mini-batch, perform a forward pass.
   Hint: consider $\mu_B = \frac{1}{n}\sum_{k=1}^{n} x_k$, $\sigma_B = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \mu_B)^2}$, $\hat{x}_i = \frac{x_i - \mu_{\mathbf{B}}}{\sigma_B}$.

2. Calculate $\frac{\partial \mathcal{L}(\hat{\mathbf{x}})}{\partial \gamma}$, $\frac{\partial \mathcal{L}(\hat{\mathbf{x}})}{\partial \beta}$ and update learnable parameters using learning rate $\alpha = 0.3$.

2 Solution

1. The computational graph is presented on Figure 1, values from the forward pass are also presented there.

2. Calculate

$$\frac{\partial \mathcal{L}(\mathbf{y})}{\partial \gamma} = \frac{\partial \mathcal{L}(\mathbf{y})}{\partial y_i}\,\frac{\partial y_i}{\partial \gamma} = 2\,y_i\,\hat{x}_i$$
$$\frac{\partial \mathcal{L}(\mathbf{y})}{\partial \beta} = \frac{\partial \mathcal{L}(\mathbf{y})}{\partial y_i}\,\frac{\partial y_i}{\partial \beta} = 2\,y_i.$$

   Now update parameters for i $\in \{1, 2\}$, following formula

$$\gamma_{\text{new}} = \gamma - \alpha\,\frac{\partial \mathcal{L}(\mathbf{y})}{\partial \gamma}$$
$$\beta_{\text{new}} = \beta - \alpha\,\frac{\partial \mathcal{L}(\mathbf{y})}{\partial \beta}.$$

   The requested update is

$$\begin{aligned} i = 1: &\quad \gamma = 4 - 0.3 \cdot 2 \cdot (-6) \cdot (-1) = 0.4, \quad \beta = -2 - 0.3 \cdot 2 \cdot (-6) = 1.6 \\ i = 2: &\quad \gamma = 0.4 - 0.3 \cdot 2 \cdot 2 \cdot 1 = -0.8, \quad \beta = 1.6 - 0.3 \cdot 2 \cdot 2 = 0.4. \end{aligned}$$

   Therefore after update, learnable parameters are $\gamma = -0.8$, $\beta = 0.4$.

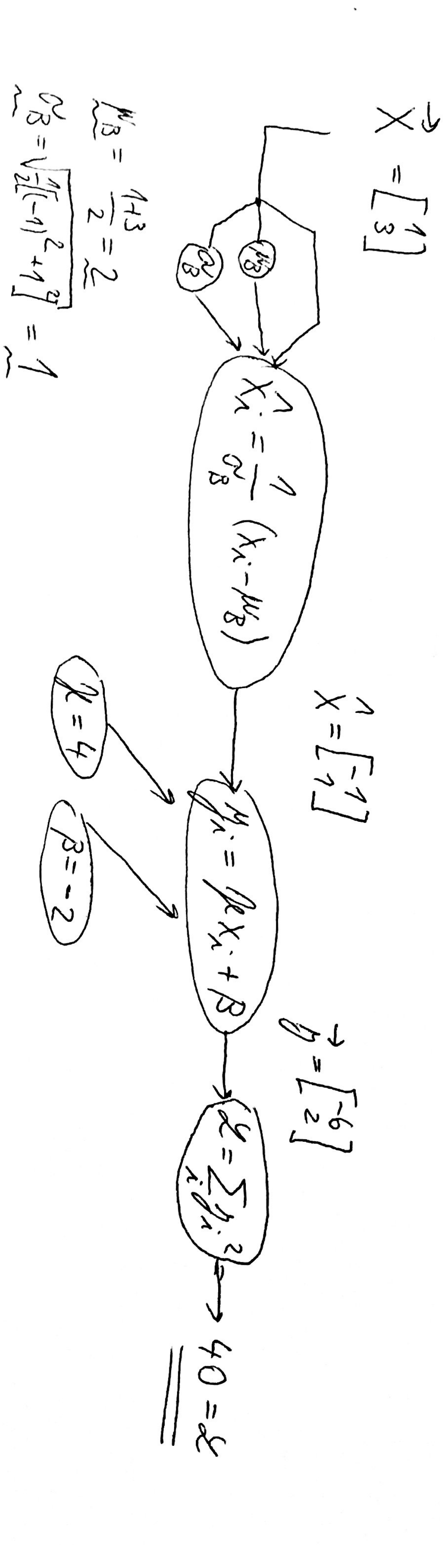

Figure 1: Computational graph - solution