Deep Learning

Exercise 2: Linear Regression via Gradient Descent

Instructor: Manuel Günther
Email: guenther@ifi.uzh.ch
Office: AND 2.54

Friday, March 5, 2020

Outline

Outline

Task 1 – Gradient Descent

Loss Function

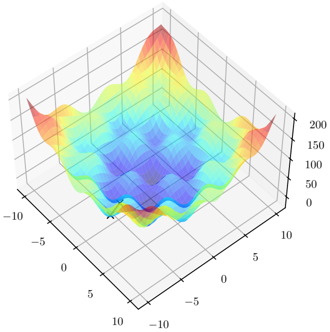
The loss function is given as: $\mathcal{J}_{\vec{w}} = w_1^2 + w_2^2 + 30 \cdot \sin(w_1) \cdot \sin(w_2)$

Task

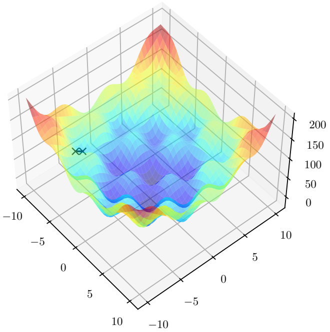
1. Compute the gradient for the loss: $\nabla_{\vec{w}}\mathcal{J}$ for a given $\vec{w}$
2. Implement gradient descent for an initial $\vec{w}$ and a given η
   - $\rightarrow$ Define an appropriate stopping criterion
   - $\rightarrow$ Return the optimized weight vector
3. Start gradient descent with several initial $\vec{w}$ and different η
   - $\rightarrow$ What and where is the minimum loss, what is an appropriate learning rate?
4. Plot the error surface for the range $w_1, w_2 \in [-10, 10]$
5. Plot start and end locations into the error surface

Taks 1 – Gradient Descent

Example Surface Plot 1



Example Surface Plot 2



Outline

Task 2 – Linear Regression via Gradient Descent

Task

1. Create 1D noisy linear training data samples:
   $X = \{(x^{[n]}, t^{[n]}) \mid 1 \leq n \leq 100\}$ with $t^{[n]} = a \cdot x^{[n]} + b +$ noise
2. Define a linear unit $y = \vec{w}^{\mathrm{T}}\vec{x}$ and a loss function $\mathcal{J}_{\vec{w}}(X)$
3. Implement a function to compute the gradient $\nabla_{\vec{w}}\mathcal{J}$
4. Initialize the weights $\vec{w}$ randomly, and choose a learning rate η
5. Implement gradient descent with appropriate termination criterion
6. Plot data points X, the initial line $a \cdot x + b$ and the learned line $y(x)$

Additional Task

1. Implement an adaptive learning rate strategy
2. Measure the speedup in terms of required epochs

Task 2 – Linear Regression via Gradient Descent

Linear Unit

$$y = \vec{w}^{\mathrm{T}}\vec{x} = w_0 + w_1 \cdot x$$

Loss Function

$$\mathcal{J}_{\vec{w}}(X) = \frac{1}{N}\sum_{n=1}^{N}\left(y^{[n]} - t^{[n]}\right)^2$$

Derivative of Loss

$$\nabla_{\vec{w}}\mathcal{J} = \begin{pmatrix} \frac{1}{N}\sum_{n=1}^{N}\left(y^{[n]} - t^{[n]}\right) \\ \frac{1}{N}\sum_{n=1}^{N}\left(y^{[n]} - t^{[n]}\right) \cdot x^{[n]} \end{pmatrix}$$

Example of Optimal Regression