Last modification: Saturday 13th June, 2020, 13:18

Warm gradient descent

Hor Dashti-Naserabadi and Matthieu Sarkis

Abstract

Random ideas

Contents

2010 Mathematics Subject Classification
Keywords:

1. List of ideas

- Generalization of mini-batch stochastic gradient descent: variable temperature (size of the mini-batch).
- Consider the mean-squared-error cost function:

$$\mathrm{MSE}(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x}^{(\mathbf{i})} - \mathbf{y}^{(\mathbf{i})}\right)^2 \tag{1}$$

  The gradient is given by:

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}}\mathrm{MSE}(\boldsymbol{\theta}) = \frac{2}{m}\,\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y}\right) \tag{2}$$

Why not then generalize the previous idea by replacing in (2) the training data matrix $\boldsymbol{X}$ by a submatrix, different at each iteration of the gradient descent. One can choose the features randomly, and adjust the 'temperature' at each step of the gradient descent.

- Polynomial data
- Non-convex loss function
- Quantum tunneling
- constant learning rate η
- $T_n = m\exp(-n)\cos(n)$ for the temperature profile
- Classification

Acknowledgements

References

Hor Dashti-Naserabadi hdashti@kias.re.kr
Korea Institute for Advanced Study 85 Hoegiro Dongdaemun-gu, Seoul 02455, Republic of Korea

Matthieu Sarkis sarkis@kias.re.kr
Korea Institute for Advanced Study 85 Hoegiro Dongdaemun-gu, Seoul 02455, Republic of Korea