(Very) Basics on Optimization

Optimization:

$$\boldsymbol{W}_o = \arg min/max_{\boldsymbol{W}} \; f(\boldsymbol{W})$$

How to solve the unconstraint optimization problem?

Solve: $$\frac{df(\boldsymbol{W})}{d\boldsymbol{W}} = \mathbf{0}$$

Gradient Descent: $$\boldsymbol{W}^{(t)} = \boldsymbol{W}^{(t-1)} - \lambda^{(t-1)} \frac{df(\boldsymbol{W})}{d\boldsymbol{W}} \Big|_{\boldsymbol{W}=\boldsymbol{W}^{(t-1)}}$$

A simple unconstrained example

Example:

$$\boldsymbol{W}_o = \arg min_{\boldsymbol{W}} \; f(\boldsymbol{W}) = ||\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}||_F^2$$

Solve:

$$\frac{df(\boldsymbol{W})}{d\boldsymbol{W}} = \frac{d||\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}||_F^2}{d\boldsymbol{W}} = \mathbf{0}$$

A simple unconstrained example

$$||\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}||_F^2 = tr[(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H})^T(\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H})]$$

$$= tr[\boldsymbol{X}^T\boldsymbol{X} - (\boldsymbol{W}\boldsymbol{H})^T\boldsymbol{X} - \boldsymbol{X}^T(\boldsymbol{W}\boldsymbol{H}) + (\boldsymbol{W}\boldsymbol{H})^T\boldsymbol{W}\boldsymbol{H}]$$

$$= tr[\boldsymbol{X}^T\boldsymbol{X}] - 2tr[(\boldsymbol{W}\boldsymbol{H})^T\boldsymbol{X}] + tr[\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T\boldsymbol{W}^T]$$

$$= tr[\boldsymbol{X}^T\boldsymbol{X}] - 2tr[\boldsymbol{X}\boldsymbol{H}^T\boldsymbol{W}^T] + tr[\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T\boldsymbol{W}^T]$$

$$\frac{dtr[\boldsymbol{X}\boldsymbol{H}^T\boldsymbol{W}^T]}{d\boldsymbol{W}} = \boldsymbol{X}\boldsymbol{H}^T \qquad \frac{dtr[\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T\boldsymbol{W}^T]}{d\boldsymbol{W}} = \boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T$$

$$\frac{d||\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}||_F^2}{d\boldsymbol{W}} = -2\boldsymbol{X}\boldsymbol{H}^T + 2\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^T = \boldsymbol{0} \Rightarrow \boldsymbol{W} = \boldsymbol{X}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{H}^T)^{-1}$$

A simple unconstrained example

Example:

$$\boldsymbol{W}_o = \arg min_{\boldsymbol{W}} \; f(\boldsymbol{W}) = ||\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}||_F^2$$

Gradient Descent:

$$\boldsymbol{W}^{(t)} = \boldsymbol{W}^{(t-1)} - \lambda^{(t-1)} \frac{df(\boldsymbol{W})}{d\boldsymbol{W}} |_{\boldsymbol{W}=\boldsymbol{W}^{(t-1)}}$$

$$\boldsymbol{W}^{(t)} = \boldsymbol{W}^{(t-1)} - \lambda^{(t-1)}(-2\boldsymbol{X}\boldsymbol{H}^T + 2\boldsymbol{W}^{(t-1)}\boldsymbol{H}\boldsymbol{H}^T)$$

Optimization with constraints

Optimization problem:

$$\boldsymbol{w}_o = \arg min/max_{\boldsymbol{w}} \ f(\boldsymbol{w})$$

$$subject\ to\ \boldsymbol{g}(\boldsymbol{w}) = \boldsymbol{0}$$

How to solve the constraint optimization problem?

A simple constrained example

Solve:

$$\boldsymbol{w}_o = \arg max_{\boldsymbol{w}} \boldsymbol{w}^T \boldsymbol{A} \boldsymbol{w}$$

$$subject\ to\ \boldsymbol{w}^T \boldsymbol{w} - 1 = 0$$

Formulate the Lagrangrian :

$$L(\boldsymbol{w}, \lambda) = \boldsymbol{w}^T \boldsymbol{A} \boldsymbol{w} - \lambda(\boldsymbol{w}^T \boldsymbol{w} - 1)$$

Solve:

$$\frac{dL(\boldsymbol{w}, \lambda)}{d\boldsymbol{W}} = 0 \Rightarrow \boldsymbol{A}\boldsymbol{w} = \lambda \boldsymbol{w}$$