5. Optimization

Dienstag, 20. November 2018 11:00

1) $\forall \vec{x},\vec{y} \in X \; \forall \lambda \in [0,1]: \; \lambda f(\vec{x}) + (1-\lambda) f(\vec{y}) \geq f(\lambda\vec{x} + (1-\lambda)\vec{y})$

X: convex set

strictly $>$

concave $\leq$, $<$

2) f diff.

convex $\Leftrightarrow \forall \vec{x},\vec{y} \in X: \; f(\vec{y}) \geq f(\vec{x}) + (\vec{y}-\vec{x})^T \vec{\nabla} f(\vec{x})$

3) f 2^{nd} diff

convex $\Leftrightarrow$ 2^{nd} der. (Hessian) ≥ 0 (pos. semidef.)

4) convexity preserving op. (sum, max, $c \cdot f$ $(c \geq 0)$, $f(A\vec{x}+\vec{b})$, $m(f(\vec{x}))$)

m: convex, non-decreasing

Problem 1: Show that affine functions of the form $\boldsymbol{w}^T\boldsymbol{x} + b$ are both convex and concave.

$$\lambda f(\vec{x}) + (1-\lambda) f(\vec{y}) \left[\overset{!}{=} f(\lambda\vec{x} + (1-\lambda)\vec{y})\right]$$

$$= \lambda(\vec{w}^T\vec{x} + b) + (1-\lambda)(\vec{w}^T\vec{y} + b) =$$

$$= \lambda\vec{w}^T\vec{x} + (1-\lambda)\vec{w}^T\vec{y} + (\lambda + 1 - \lambda) b =$$

$$= \vec{w}^T(\lambda\vec{x} + (1-\lambda)\vec{y}) + b =$$

$$= f(\lambda\vec{x} + (1-\lambda)\vec{y})$$

Problem 2: Show that a twice differentiable function $f(\boldsymbol{x})$ with a convex domain is convex if and only if its Hessian or second derivative is positive semidefinite: $\nabla^2 f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathrm{dom}(f)$.

1D: "$\Rightarrow$" $\underline{f'(x)(y-x) \leq f(y) - f(x)}$

$f'(y)(x-y) \leq f(x) - f(y) \Leftrightarrow f(y) - f(x) \leq \underline{f'(y)(y-x)}$

$$f'(y)(x-y) \leq f(x)-f(y) \Leftrightarrow f(y)-f(x) \leq f'(y)(y-x)$$

$$(f'(y)-f'(x))(y-x) \geq 0 \quad \Big| : (y-x)^2$$

$$\frac{f'(y)-f'(x)}{y-x} \geq 0$$

$$y \to x \;\Rightarrow\; f''(x) \geq 0$$

"$\Leftarrow$" $f''(z) \geq 0 \quad x,y \in dom(f)$

$$0 \leq \int_x^y f''(z)(y-z)\,dz \underset{p.I.}{=} \left[f'(z)(y-z)\right]_{z=x}^{y} + \int_x^y f'(z)\,dz =$$

$$= -f'(x)(y-x) + f(y) - f(x)$$

$$\Leftrightarrow f(y) \geq f'(x)(y-x) + f(y)$$

$n>1$: $\quad g(t) = f(\vec{x}_0 + t\vec{v}) \quad \forall \vec{x}_0 + t\vec{v} \in dom(f)$, convex for all $\vec{x}_0, t, \vec{v} \Leftrightarrow f$ convex

$$g''(t) = \vec{v}^T \vec{\nabla}^2 f(\vec{x}_0 + t\vec{v})\vec{v} \geq 0 \underset{1D}{\Leftrightarrow} g(t) \text{ convex}$$

Problem 3: Prove that the objective function of logistic regression

$$E(\boldsymbol{w}) = -\ln p(\boldsymbol{y} \mid \boldsymbol{w}, \boldsymbol{X}) = -\sum_{i=1}^{N} \big(\overset{\geq 0}{y_i \ln \sigma(\boldsymbol{w}^T \boldsymbol{x}_i)} + \overset{\geq 0}{(1-y_i)\ln(1-\sigma(\boldsymbol{w}^T\boldsymbol{x}_i))}\big) \tag{1}$$

is convex. What is the benefit of having a convex function for optimization?

$-\ln \sigma(\vec{w}^T \vec{x}_i)$ ✓ $\qquad -\ln(1-\sigma(\vec{w}^T\vec{x}_i))$ ✓

$$\partial_x \sigma(x) = \partial_x \frac{1}{1+e^{-x}} = \frac{+e^{-x}}{(1+e^{-x})^2} = \sigma(x)\frac{e^{-x}}{1+e^{-x}} =$$

$$= \sigma(x)\left(1 - 1 + \frac{e^{-x}}{1+e^{-x}}\right) = \sigma(x)\left(1 + \frac{-1 \cancel{-e^{-x}} \cancel{+e^{-x}}}{1+e^{-x}}\right) =$$

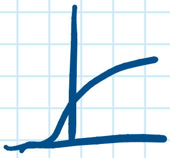
$$= \sigma(x)(1-\sigma(x))$$

$$\vec{\nabla}_{\vec{w}}^2\left(-\ln \sigma(\vec{w}^T\vec{x}_i)\right) = \vec{\nabla}_{\vec{w}}\left(-\vec{x}_i \frac{\cancel{\sigma(\vec{w}^T\vec{x}_i)}(1-\sigma(\vec{w}^T\vec{x}_i))}{\cancel{\sigma(\vec{w}^T\vec{x}_i)}}\right) =$$

$$= \vec{\nabla}_{\vec{w}}\left[-\vec{x}_i(1-\sigma(\vec{w}^T\vec{x}_i))\right] = \vec{x}_i\vec{x}_i^T\sigma(\vec{w}^T\vec{x}_i)(1-\sigma(\vec{w}^T\vec{x}_i))$$

$$= \vec\nabla_{\vec w}\left[-\vec x_i\left(1-\sigma(\vec w^T\vec x_i)\right)\right] = \vec x_i\vec x_i^T\sigma(\vec w^T\vec x_i)\left(1-\sigma(\vec w^T\vec x_i)\right)$$

$$\forall \vec z\in\mathbb{R}^n:\quad \underbrace{\vec z^T\left[\vec x_i\right.}\underbrace{\left.\vec x_i^T\sigma(\vec w^T\vec x_i)\left(1-\sigma(\vec w^T\vec x_i)\right)\right]\vec z}\overset{?}{\geq} 0$$



$$\underbrace{(\vec x_i^T\vec z)^2}_{\geq 0}\,\underbrace{\sigma(\vec w^T\vec x_i)}_{\geq 0}\underbrace{\left(1-\sigma(\vec w^T\vec x_i)\right)}_{\geq 0}\overset{?}{\geq} 0$$

Right: $-\ln\left(1-\sigma(\vec w^T\vec x_i)\right) = -\ln\left(1-\frac{1}{1+e^{-\vec w^T\vec x_i}}\right) = -\ln\left(\frac{\cancel{1}+e^{-\vec w^T\vec x_i}\cancel{-1}}{1+e^{-\vec w^T\vec x_i}}\right) =$

$$= \cancel{-\ln}\left(\cancel{e}^{+\vec w^T\vec x_i}\right) - \ln\left(\underbrace{\frac{1}{1+e^{-\vec w^T\vec x_i}}}_{\sigma}\right) = \underbrace{\underbrace{\vec w^T\vec x_i}_{\substack{\text{linear}\\\text{convex}}}\underbrace{-\ln\sigma(\vec w^T\vec x_i)}_{\text{convex}}}_{\text{convex}}$$

Problem 4: Discuss the following topics:

- Condition number
- Consistency, convergence, stability
- Stiffness

Condition number: rel. output error from rel. input error (prop. of problem)

Consistency: Local discretization error $\ell(\delta t)\to 0$ as $\delta t\to 0$

Convergence: Global discretization error $e(\delta t)\to 0$ as $\delta t\to 0$

Stability: Doesn't magnify approximation errors

Stability + Consistency = Convergence

Stiffness: Local property of algorithm's solution imposes high resolution