Lec 18

Thursday, November 14, 2019 10:59

Recap: last time: Various first-order optimization algos

that is, algos that use gradient info

i.e. only <u>first</u> deriv

Fitting NNs w/ Gradient Descent: <u>Backpropagation</u>

$$R(\theta) = \sum_i \sum_k L(y_{ik}, f_k(x_i ; \theta))$$

e.g. regression $L(y, \hat{y}) = (y - \hat{y})^2$

classification $L(y, \hat{p}) = -y \log \hat{p}$

$\theta = \{$all the weights $\alpha_{0m}, \alpha_m, \beta_{0k}, \beta_k$ parameterizing our vanilla neural net$\}$

Want θ to min $R(\theta)$

Use first-order methods

So need to compute $\nabla R(\theta)$

The key: Chain rule

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x}$$

Focus on $L(y, \hat{y}) = (y - \hat{y})^2$

&

$$= \quad f_k(x_i, \theta) = \beta_k z_i + \beta_0 \quad (g_k = \text{identity})$$

$$R(\theta) = \sum_i \underbrace{\sum_k \left(y_{ik} - f_k(x_i; \theta) \right)^2}_{R_i(\theta)}$$

Want $\dfrac{\partial R_i}{\partial \beta_{km}}$, $\dfrac{\partial R_i}{\partial \alpha_{ml}}$ for all k, m, l

$$\frac{\partial R_i}{\partial \beta_{km}} = \sum_{k'} \frac{\partial R_i}{\partial f_{k'}(x_i)} \cdot \underbrace{\frac{\partial f_{k'}(x_i)}{\partial \beta_{km}}}_{=0 \text{ whenever } k \neq k'}$$

$$= \frac{\partial R_i}{\partial f_k(x_i)} \cdot \frac{\partial f_k(x_i)}{\partial \beta_{km}}$$

$$= \underbrace{-2\left(y_{ik} - f_k(x_i) \right)}_{\delta_{ki}} z_{mi}$$

convention $z_{0i} = 1$

$$\frac{\partial R_i}{\partial \alpha_{ml}} = \sum_k \frac{\partial R_i}{\partial f_k(x_i)} \sum_{m'} \frac{\partial f_k(x_i)}{\partial z_{m'l}} \underbrace{\frac{\partial z_{m'l}}{\partial \alpha_{ml}}}_{=0 \text{ whenever } m \neq m'}$$

$$= \sum_k \frac{\partial R_i}{\partial f_k(x_i)} \frac{\partial f_k(x_i)}{\partial z_{ml}} \frac{\partial z_{ml}}{\partial \alpha_{ml}}$$

$$= \underbrace{\sum_k \left(-2\left(y_{ik} - f_k(x_i) \right) \right) \beta_{km} \, \sigma'(z_m)}\, x_{il}$$

$s_{mi} = \sum_k \delta_{ki} \cdot \beta_{km} \sigma'(z_{mi})$ back prop

$$\frac{\partial R_i}{\partial \beta_{km}} = \underline{\delta_{ki}}\, z_{mi} \qquad \frac{\partial R_i}{\partial \alpha_{ml}} = \underline{s_{mi}}\, x_{il}$$

"errors"