TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Dropout

Dropout

Dropout can be viewed as an ensemble method.

To draw a model from the ensemble we randomly select a mask μ with

$$\begin{cases} \mu_i = 0 \text{ with probability } \alpha \\ \\ \mu_i = 1 \text{ with probability } 1 - \alpha \end{cases}$$

Then we use the model $(\Phi,\ \mu)$ with weight layers defined by

$$y_i = \text{Relu}\left(\sum_j W_{i,j}\ \mu_j x_j\right)$$

Dropout Training

Repeat:

- Select a random dropout mask μ
- Φ -= $\nabla_\Phi \mathcal{L}(\Phi, \mu)$

Backpropagation must use the same mask μ used in the forward computation.

Test Time Scaling

At train time we have

$$y_i = \text{Relu}\left(\sum_j W_{i,j}\ \mu_j x_j\right)$$

At test time we have

$$y_i = \text{Relu}\left((1-\alpha)\ \sum_j W_{i,j}\ x_j\right)$$

At test time we use the "average network".

Dropout for Least Squares Regression

Consider simple least square regression

$$\begin{aligned}
\Phi^* &= \underset{\Phi}{\operatorname{argmin}} \;\; \mathrm{E}_{(x,y)} \, E_\mu \left(y - \Phi \cdot (\mu \odot x)\right)^2 \\
&= \mathrm{E}\left[(\mu \odot x)(\mu \odot x)^\top\right]^{-1} \; \mathrm{E}\left[y(\mu \odot x)\right] \\
&= \underset{\Phi}{\operatorname{argmin}} \;\; \mathrm{E}_{(x,y)}(y - (1-\alpha)\Phi \cdot x)^2 + \sum_i \frac{1}{2}(\alpha - \alpha^2)\mathrm{E}\left[x_i^2\right] \Phi_i^2
\end{aligned}$$

In this case dropout is equivalent to a form of L_2 regularization — see Wager et al. (2013).

A Dropout Bound

$$
\begin{aligned}
KL(Q_{\alpha,\Phi},\ Q_{\alpha,0}) &= E_{\mu\sim P_\alpha,\epsilon\sim\mathcal{N}(0,1)^d}\ \ln\frac{P_\alpha(\mu)e^{-\frac{1}{2}||\mu\odot\epsilon||^2}}{P_\alpha(\mu)e^{-\frac{1}{2}||\mu\odot(\Phi+\epsilon)||^2}} \\
&= E_{\mu\sim P_\alpha}\ \frac{1}{2}||\mu\odot\Phi||^2 \\
&= \frac{1-\alpha}{2}||\Phi||^2
\end{aligned}
$$

$$
L(Q_{\alpha,\Phi}) \leq \frac{1}{1-\frac{1}{2\lambda}}\left(\hat{L}(Q_{\alpha,\Phi}) + \frac{\lambda L_{\max}}{N}\left(\frac{1-\alpha}{2}||\Phi||^2 + \ln\frac{1}{\delta}\right)\right)
$$

END