Regularization in neural networks

Definition

A process of introducing additional information in order to solve an ill-posed problem or to prevent overfitting.
It penalizes the loss function by adding a multiple of an L1 (Lasso) or an L2 (Ridge) norm of your weights vector w.

Why

Solve the overfitting problem.

Formulation

Cost function over *m* training examples

$$\frac{1}{m}L(\widehat{y}^{(i)},y^{(i)}) + \lambda * R(w)$$

Hyperparameters

- λ: the regularization parameter

Variations

L1 regulatization

Adds the absolute values of the model's coefficients as the penalty term.

$$R(w) = \frac{1}{m}\sum_{l=1}^{L}|w^{[l]}|$$

L2 regulatization

Adds the squared magnitude of the model's coefficients as the penalty term.

$$R(w) = \frac{1}{2m}\sum_{l=1}^{L}\|w^{[l]}\|^2 = \frac{1}{2m}\sum_{l=1}^{L}\sum_{i=i}^{n^{[l-1]}}\sum_{j=1}^{n^{[l]}}(w_{ij}^{[l]})^2$$

New formula for weight update

$$W^{[l]} = W^{[l]} - \alpha * dW^{[l]} = W^{[l]} - \alpha * (amount\ from\ backprob + \frac{\lambda}{m}W^{[l]})$$

Elastic net (L1 + L2)

Adds both L1 and L2 penalities.

$$R(w) = \sum_{l=1}^{L}\sum_{i=1}^{n^{[l-1]}}\sum_{j=1}^{n^{[l]}}\left(\beta * (w_{i,j}^{[l]})^2 + |w_{i,j}^{[l]}|\right)$$