Artificial neural networks (ANNs)

$$F_\theta : \mathcal{X} \longrightarrow \mathcal{Y}$$

θ: tunable parameters of the ANN (weights & bias)

Stochastic Gradient Descent (SGD)

- Suppose we have an ANN F_θ. Then the update rule has the following form:

$$\theta^{(k+1)} \leftarrow \theta^{(k)} - \eta \nabla_\theta \ell(\theta^{(k)})$$

η: learning rate (fixed, $\ll 1$)

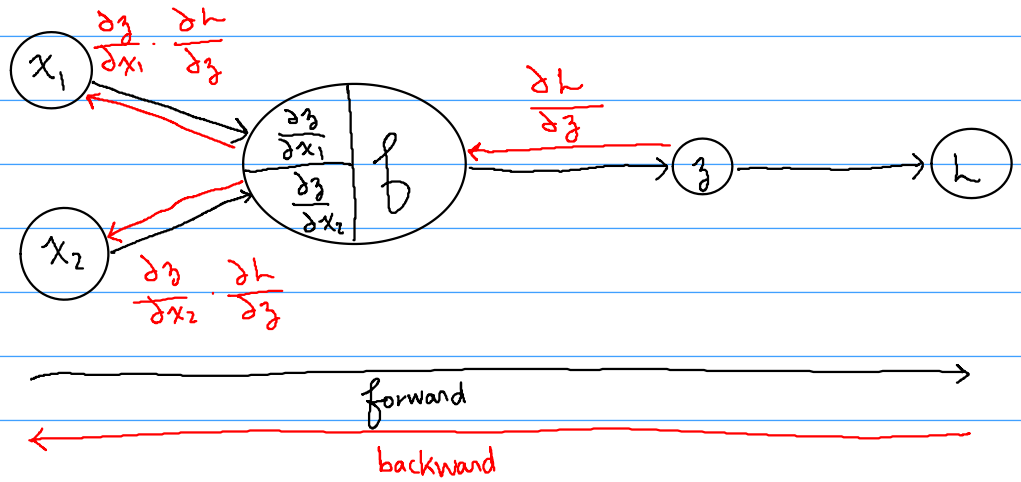
$\nabla_\theta \ell(\theta^{(k)})$

We use mini-batches of data points to compute the gradient ($M = 1, 100, 1000$).

- There exist several update rules, e.g. Adam, momentum, etc.

Backpropagation

- Suppose $x_1, x_2 \longmapsto f(x_1, x_2) = z \longmapsto L(z)$.



Classification / Regression using ANNs

ANNs can be used in other classification or regression problems. Indeed,

- alternative to multi-class logistic regression.
- Poisson regression
- etc.