1. Introduction

Y - the variable that we predict.

Feature(x) the variable in the data vector. Types:

1. Numerical
2. Categorical
   - Ordinal
   - Nominal

Hyperparam meta parameter for model. Model do not learn it.

1.1. Supervised

Solves regression and classification tasks.

$$X \longrightarrow F \longrightarrow y$$

Regression model predicts continuous values.
Classification model predicts categorical values.

1.2. Unsupervised

$$X \longrightarrow F \longrightarrow X'$$

1.3. Reinforement learning

TODO

2. Optimisation and loss function

2.1. Gradient descent

Gradient(∇f) defines direction and rate of fastest increase of scalar-valued differentiable function f.

Example for gradient in cartesian coordinate system f:

$$\nabla f = \frac{\partial f}{\partial x} i + \frac{\partial f}{\partial y} i + \frac{\partial f}{\partial z} k$$

Gradient descent iterative optimization algorithm of the first order to find the local minimum of the function.

Stop criteria for the gradient descent can be a threshold for the gradient value.

2.2. Optimization

Optimisation target minimize loss function.

Simple example of the loss function is a MSE.
Mean squared error(MSE) measures the average of squeared errors.

$$\text{MSE} = \frac{1}{N} \sum_{(x,y) \in D} \left(y - \text{prediction}(x)\right)^2$$

Iteration step for model: $x^{i+1} = x^i - h \frac{\partial f}{\partial x^i}$

Where $\frac{\partial f}{\partial x^i}$ is a *gradient*.

Possible data slices for 1 iteration there model updates params:

- simple - 1 full dataset
- stochastic - 1 record
- mini-batch - batch of random examples(e.g. 10-1000)

2.3. Regularization

TBD