

## 01. Intro

### What is

1. Supervised Learning - Classification/Regression
2. Unsupervised - Kmeans, PCA
3. Reinforcement - Giving rewards and punishment

### Linear regression

- Data - tuples of points  $\{(x_n, t_n)\}_{0 \rightarrow N}$
- Model -  $y = w_0 + xw_1$
- Loss - Mean square error  $\frac{1}{2} \sum_{n=1}^N (y_n - t_n)^2$
- Derivative to find lowest point to find  $w$
- $t_n = [x_n][w_n]^T$

### Non-Linear regression

- Basis factor expansion  $x \rightarrow [x, x^2, x^3 \dots, x^P]$
- Find weights for each polynomial/exponential/..
- Note that minimization of error function has a unique solution
  - Unique weight that optimises the model

### Regularization

#### Ridge regression solution

- Add regularizer  $(\frac{\lambda}{2} ||w||^2)$  to error function
  - Lambda is a constant
  - smaller = error matters more
  - larger = penalise weights more

## 02. Regression II

- $t_n = y(x_n, w) + \epsilon$ 
  - $\epsilon$  is random "noise" modelled via some distribution
- Assumption is that each point is IID
  - Independently identically distributed variables
- MLE is a generalised model estimator given a set of data and predicted distribution

### Examples

- MSE is a loss function for MLE assuming data is normally distributed
- Variance of gaussian known but mean is unknown
- Model used to derive mean and minimise loss/maximum likelihood

### Bayesian Linear Regression

- $posterior = \frac{likelihood \times prior}{evidence} = \frac{p(y|w)p(w)}{p(y)}$
- Repeatedly updating model based on additional data points
- Initial estimator (prior) that can be modelled in different distributions
  - By CLM, posterior distribution becomes independent to prior
  - Prior models normal distribution
- Maximum-a-Posteriori estimation (MAP)
  - For given prior, after observing data, how to update distr of params?
  - Adds "regularisation" component to normally distributed datasets