Day 3: Hyperparameter Tuning.

Parameters which define the model architecture are referred to as hyperparameters.

The process of searching an ideal model architecture is referred to as hyperparameter tuning.

* The process for learning parameter values.

• Model based learning

| use input data | :→ |

$$\begin{bmatrix} x_{1\cdot0} & x_{1\cdot1} & \cdots & x_{1\cdot n} \\ x_{2\cdot0} & x_{2\cdot1} & & x_{2\cdot n} \\ \vdots & & & \\ x_{m\cdot0} & x_{m\cdot1} & \cdots & x_{m\cdot n} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{bmatrix}$$

| To learn a set of parameters. | → | $\begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix}$.

| Yields a generalized function. | → | $f(x; \theta) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots \theta_n x_n$

| capable of predicting values or classes on new input data | → | $f(x_i; \theta) = 39.$

$$f(x_j; \theta) = 1.$$

* Hyperparameter Tuning Methods.

i) Grid Search :-

It simply builds a model for each possible combination of all the hyperparameter values provided, evaluating each model, and selecting the architecture which produces the best results.

Each model would be fit to training data and evaluated on a validation data

• Its an Exhaustive sampling of HT space and can be quite inefficient.

ii) Random Search :-

↳ Differs from Grid Search.
↳ provide a statistical distribution for each hyperparameter

iii) Bayesian optimization :

↳ Sequential model based optimization
↳ allows to use the results of previous iteration to improve our sampling method of next experiment.