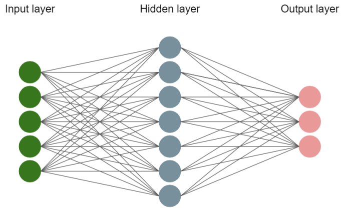
Introduction to Machine Learning

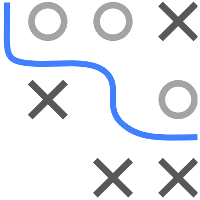
ML-Basics
Models & Parameters



Learning goals

- Understand that an ML model is simply a parametrized function
- Understand that the hypothesis space lists all admissible models
- Understand relationship between hypothesis and parameter space

WHAT IS A MODEL?

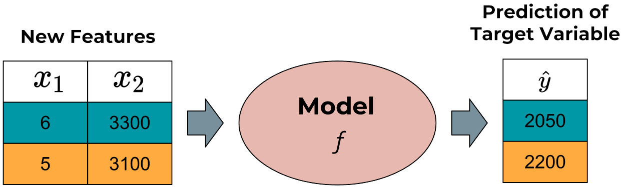


- A **model** (or **hypothesis**)

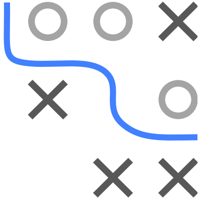
$$f : \mathcal{X} \to \mathbb{R}^g$$

  is a function that maps feature vectors to predicted target values

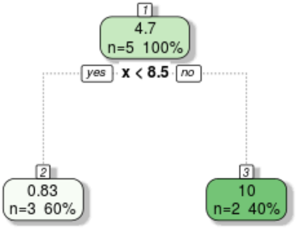
- In regression: $g = 1$; in classification, g is the number of classes, and output vectors are scores or class probabilities

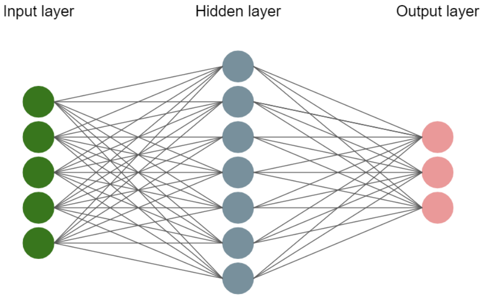


WHAT IS A MODEL?



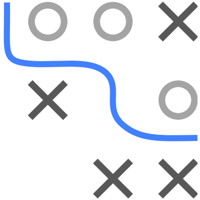
- f is meant to capture intrinsic patterns of the data, the underlying assumption being that these hold true for *all* data drawn from $\mathbb{P}_{xy}$
- Models can range from super simple (e.g., linear, tree stumps) to very complex (e.g., DL) with lots of choices





- ML requires **constraining** f to a certain type of functions

HYPOTHESIS SPACES



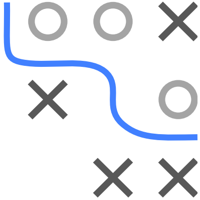
- Without restrictions on the functional family, the task of finding a “good” model among all available models is impossible to solve

- We have to determine the class of our model *a priori*, thereby narrowing down the search space. We call this a **structural prior**.

- The set of functions defining a specific model class is called a **hypothesis space** $\mathcal{H}$:

$$\mathcal{H} = \{f : f \text{ belongs to a certain functional family}\}$$

PARAMETRIZATION

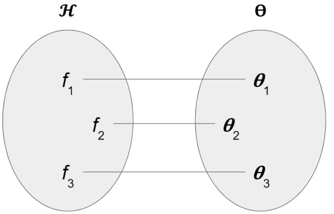


- All models within a hypothesis space share a common functional structure typically constructed as **parametrized family of functions**
- We collect all parameters in a **parameter vector** $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ from **parameter space** Θ
- They are our means of fixing a specific function from the family: once set our model is fully determined
- Therefore, we can re-write $\mathcal{H}$ as:

$$\mathcal{H} = \{f_{\boldsymbol{\theta}} : \quad f_{\boldsymbol{\theta}} \text{ belongs to a certain functional family parameterized by } \boldsymbol{\theta}\}$$

PARAMETRIZATION

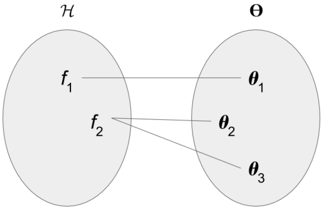
- Finding optimal model $=$ finding optimal parameters
- This allows us to operationalize our search for the best model as a search for the optimal value on a d-dimensional parameter surface



- θ might be scalar or very high-dimensional with thousands of parameters depending on the complexity of our model

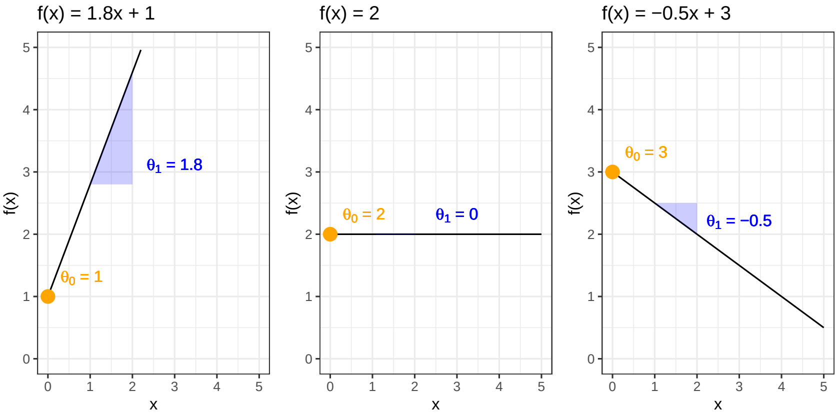
PARAMETRIZATION

- Some parameter vectors, for some model classes, encode the same function: the parameter-to-model mapping could be non-injective
- We call this a non-identifiable model
- This shall not concern us here



EXAMPLE: UNIVARIATE LINEAR FUNCTIONS

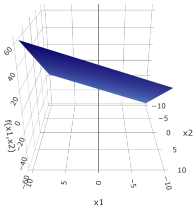
$$\mathcal{H} = \{f : f(\mathbf{x}) = \theta_0 + \theta_1 x, \boldsymbol{\theta} \in \mathbb{R}^2\}$$



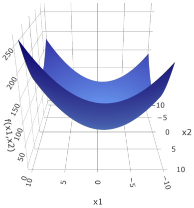
EXAMPLE: BIVARIATE QUADRATIC FUNCTIONS

$$\mathcal{H} = \{f : f(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2, \boldsymbol{\theta} \in \mathbb{R}^6\}$$

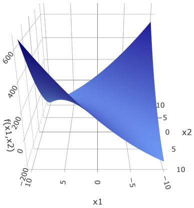
$f(x) = 3 + 2x_1 + 4x_2$



$f(x) = 3 + 2x_1 + 4x_2 +$
$+ 1x_1^2 + 1x_2^2$



$f(x) = 3 + 2x_1 + 4x_2 +$
$+ 1x_1^2 + 1x_2^2 + 4x_1 x_2$



EXAMPLE: RBF NETWORK

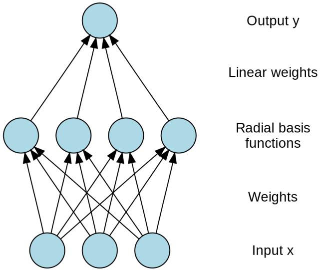
Radial basis function networks with Gaussian basis functions

$$\mathcal{H} = \left\{ f : f(\mathbf{x}) = \sum_{i=1}^{k} a_i \rho(\|\mathbf{x} - \mathbf{c}_i\|) \right\}$$

where

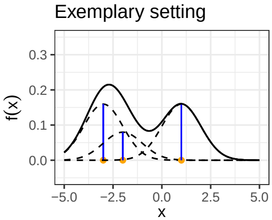
- a_i is the weight of the i-th neuron
- $\mathbf{c}_i$ its center vector and
- $\rho(\|\mathbf{x} - \mathbf{c}_i\|) = \exp(-\beta\|\mathbf{x} - \mathbf{c}_i\|^2)$ is the i-th radial basis function with bandwidth $\beta \in \mathbb{R}$

Usually number of centers k and bandwidth β need to be set in advance (so-called *hyperparameters*)

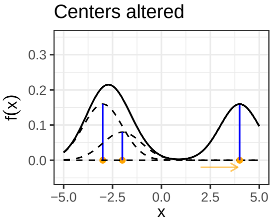


▸ Click for source

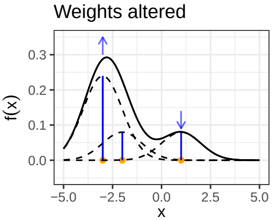
EXAMPLE: RBF NETWORK



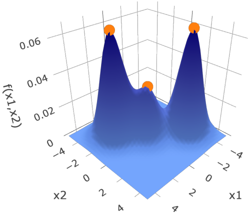
$a_1 = 0.4, a_2 = 0.2, a_3 = 0.4$
$c_1 = -3, c_2 = -2, c_3 = 1$



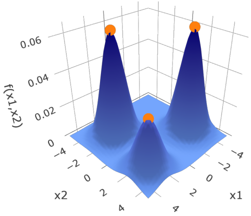
$a_1 = 0.4, a_2 = 0.2, a_3 = 0.4$
$c_1 = -3, c_2 = -2, c_3 = 4$



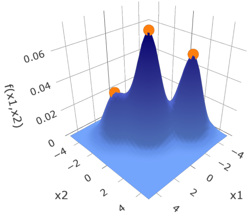
$a_1 = 0.6, a_2 = 0.2, a_3 = 0.2$
$c_1 = -3, c_2 = -2, c_3 = 1$



$a_1 = 0.4, a_2 = 0.2, a_3 = 0.4$
$c_1 = (2, -2), c_2 = (0, 0),$
$c_3 = (-3, 2)$



$a_1 = 0.4, a_2 = 0.2, a_3 = 0.4$
$c_1 = (2, -2), c_2 = (3, 3),$
$c_3 = (-3, 2)$



$a_1 = 0.2, a_2 = 0.45, a_3 = 0.35$
$c_1 = (2, -2), c_2 = (0, 0),$
$c_3 = (-3, 2)$