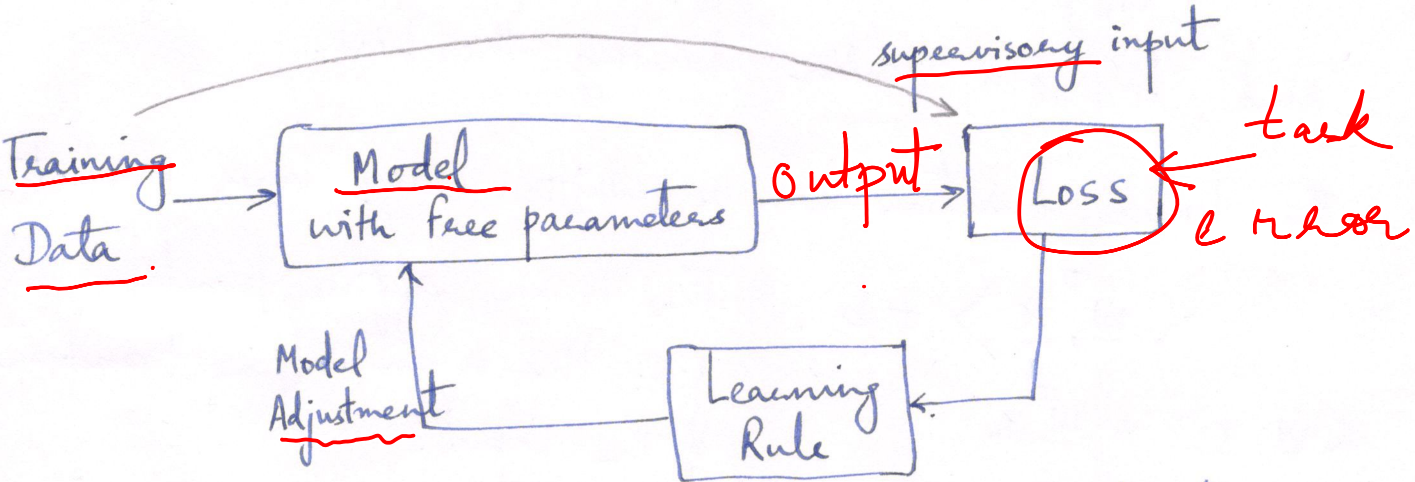
Linear Predictors (Chapter 9)



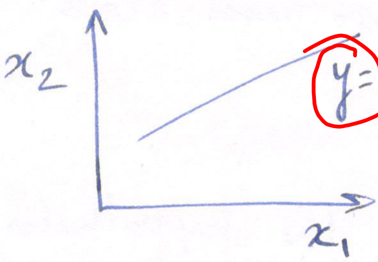
Learning rule is the Guiding Principle for learning

Linear Predictor

The model is a linear function of the inputs

→ predict classification labels (Linear classification)

→ predict some real no. $\mathbb{R}$. (Linear Regression)



$y = ax_1 + bx_2 + c = 0$

Parameters $[a, b, c]$ define the line.

These parameters can be put together to construct a vector $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$

The input can be considered as a vector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ for 2D input

Parameters which multiply with the input components x_1 & x_2 are called as weights and denoted with a vector $\underline{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

The linear function of the input is $w_1 x_1 + w_2 x_2 + b$

$\underline{w} \equiv \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ are the weight components

and $\underline{b}$ is a scalar denoting the bias.

The linear function can be written using inner product

$$\underbrace{\langle \underline{w}^T, \underline{x} \rangle}_{} \underbrace{+ b}_{}$$

This is an affine function i.e. this is an affine transform of the input.

$\langle \underline{w}, \underline{x} \rangle = \underline{w}^T \underline{x}$

The predicted value $y \equiv \langle \underline{w}, \underline{x} \rangle + b$ scales linearly with inputs x_1 and x_2.

w_1
w_2
b

$\langle \underline{w}, \underline{x} \rangle + b$ represents a family of linear functions

This is called as the hypothesis class of linear functions

A single hypothesis from this hypothesis class $\mathcal{H}$

$\mathcal{L}_d$ d: dim

is $h_{\underline{w},b}(\underline{x}) = \langle \underline{w}, \underline{x} \rangle + b$.

$h_{\underline{w},b}$

The learning problem is to find the hypothesis that best fits the training data. i.e. best approximates the underlying mapping from the input $\underline{x}$ to the label y

learning rule.

	Input features	Label ground truth
	x_1	y_1
x_i : ith input	x_2	y_2
y_i : target label for the ith input	x_3	y_3
	$\vdots$	
	x_m	y_m

$\underline{x}$: vector.
b : scalar.

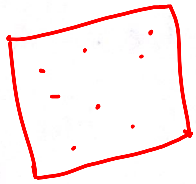
A

$$h(\underline{x}) = w_1 x_1 + w_2 x_2 + b \qquad 2d \qquad \langle \underline{w}, \underline{x} \rangle$$

$$= \underbrace{\underline{w}^T \underline{x}} + b = \underbrace{\langle \underline{w}, \underline{x} \rangle} + b.$$

In general, $\underline{x}$: d-dim vector

$\underline{w}$: d-dim vector



2d space

$$h_{\underline{w},b}. \quad h(\underline{x}) = \underbrace{\sum_{i=1}^{d} w_i x_i}_{\langle \underline{w}, \underline{x} \rangle} + b$$

Unknown parameters of the model are $\underline{w} \in \mathbb{R}^d$, $b \in \mathbb{R}$.

The training set $S \equiv \underbrace{(\underline{x}_i, y_i)_m}$.

The hypothesis (function) defined by parameters $(\underline{w}, b)$ is denoted as

$$\underbrace{h_{\underline{w},b}(\underline{x})} \longmapsto \underbrace{\langle \underline{w}, \underline{x} \rangle + b}$$ affine

The hypothesis class available for the linear predictor is the set of functions

$$\mathcal{H} \ \left\{ \underbrace{h_{\underline{w},b}(\underline{x})} ; \ \underbrace{\underline{w} \in \mathbb{R}^{d}}, \ b \in \underbrace{\mathbb{R}} \right\}$$

$\mathbb{R}$ real space

$\mathbb{R}^d$ d-dim real space.

The hypothesis class available for a linear predictor is denoted as L_d – the set of affine maps.

$$\underline{x} \longmapsto \langle \underline{w}, \underline{x} \rangle + b$$

Writing in another way

$$\underline{x} \longmapsto \langle \tilde{\underline{w}}, \tilde{\underline{x}} \rangle$$

$$\tilde{\underline{w}} \equiv \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \\ b \end{bmatrix} \qquad \tilde{\underline{x}} \equiv \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \\ 1 \end{bmatrix}$$

The last component is constant input 1