COMBINING MULTIPLE LEARNERS

- many different algorithms/learners

- NO FREE LUNCH THEOREM ⇒ no single algorithm is always the best one.

- several algorithms
- several hyperparameters
  - k-NN (3-NN, 5-NN, 7-NN, ...)
  - MLP (# of hidden nodes, activation functions)

- MAIN IDEA ⇒ DIVERSITY

(1) How do we generate base-learners that complement each other?

If they produce ~~the same~~ very similar predictions, they do not complement each other.

	f_1	f_2	...	f_K
x_1	+	−	⋮	+
x_2	+	+	⋮	+
⋮	⋮	⋮	⋮	⋮
x_N	−	−	− − −	−

$$(f_1 + f_2 + \dots + f_K) / K$$

⇓

majority voting

- if positives have the majority (+)
- if negatives have the majority (−)

(2) How do we combine the outputs of base-learners for obtaining the maximum accuracy?

Generating Diverse Learners:

(1) Different Algorithms → MCP + k-NN
"inductive bias" → MLP + k-NN + SVM + DT
→ one parametric + one nonparametric

(2) Different Hyperparameters

k-NN → k=3 (local), k=17 (global)

MLP → H=50 (simpler), H=500 (more complex)

(3) Different Input Representations "multi view" or "multimodal" learning.

different types of sensors/measurements/modalities

sensor fusion ⇒ audio + video

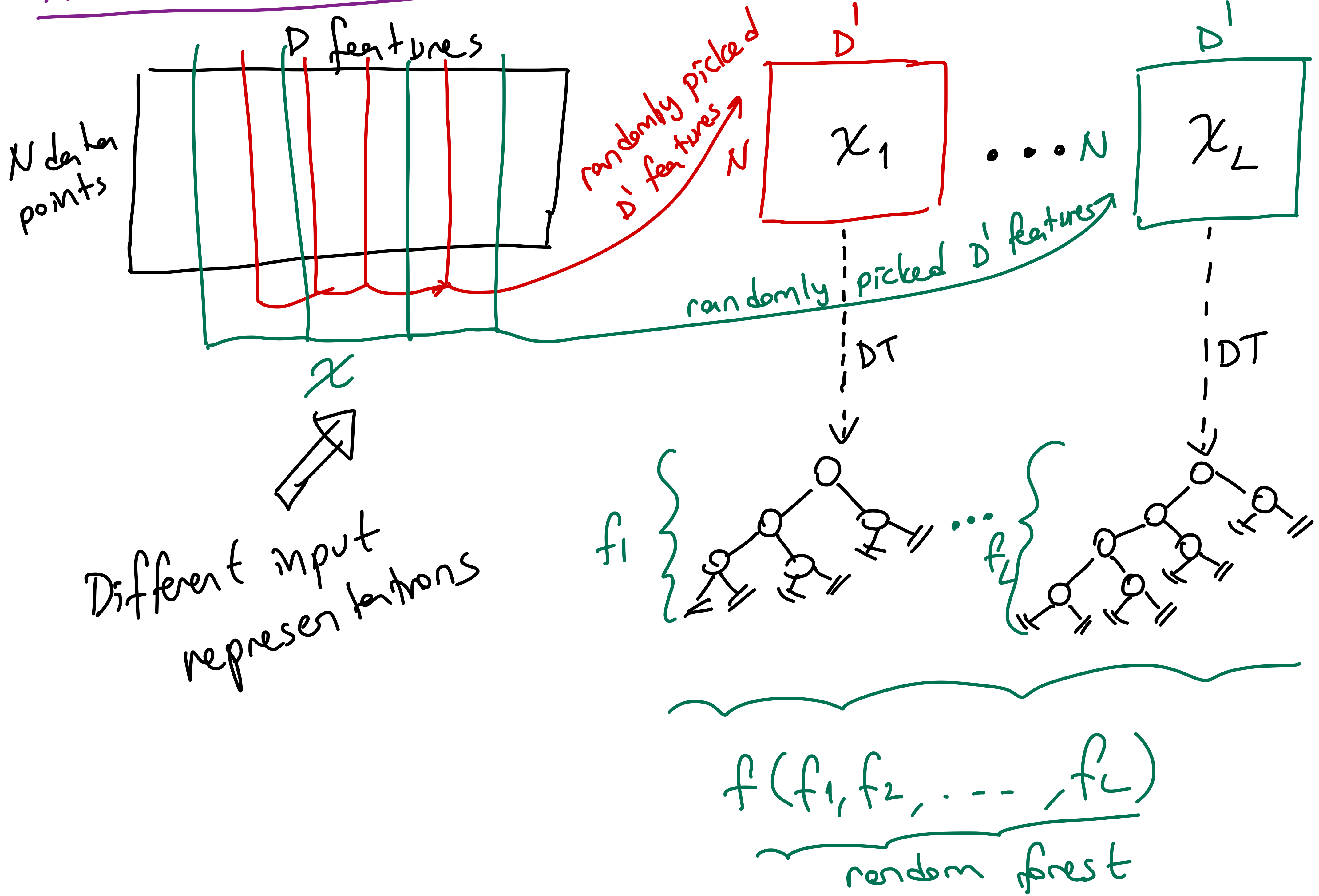
audio ↓ speech

video ↓ movements

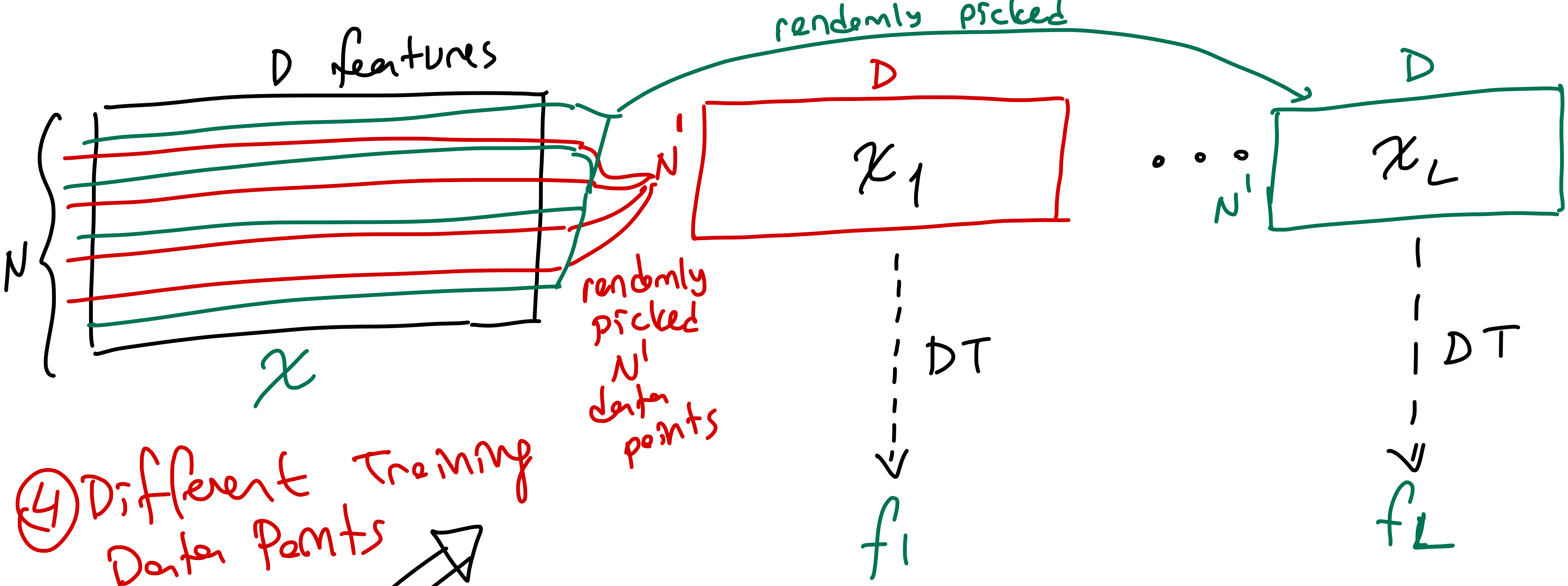
$x = [\, x_a \mid x_v \,]$ early fusion f

late fusion x_a f_a, x_v f_v

RANDOM FOREST:



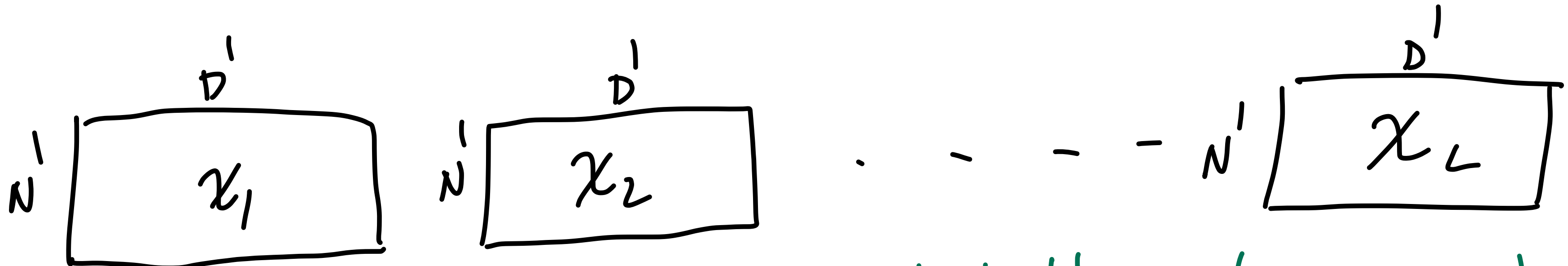
(4) Different Training Data Points



$f(f_1, f_2, \ldots f_L)$

random forest

Different training data points



randomly subset both columns and rows

Model Combination Strategies:

multiple expert combination → global combination (learner fusion)
→ local combination (learner selection)

$f_1 \quad f_2 \quad \cdots \quad f_L$

$L = \#$ of base-learners

$x_{N+1} \Rightarrow$ test data

$f_1(x_{N+1}) \quad f_2(x_{N+1}) \quad \cdots \quad f_L(x_{N+1})$

Combination

$$\boxed{w_1 f_1(x_{N+1}) + w_2 f_2(x_{N+1}) + \cdots + w_L f_L(x_{N+1})} = f$$

Majority Voting: $w_1 = 1 \quad w_2 = 1 \quad \cdots \quad w_L = 1.$

if $f(x_{N+1}) > 0 \Rightarrow +$

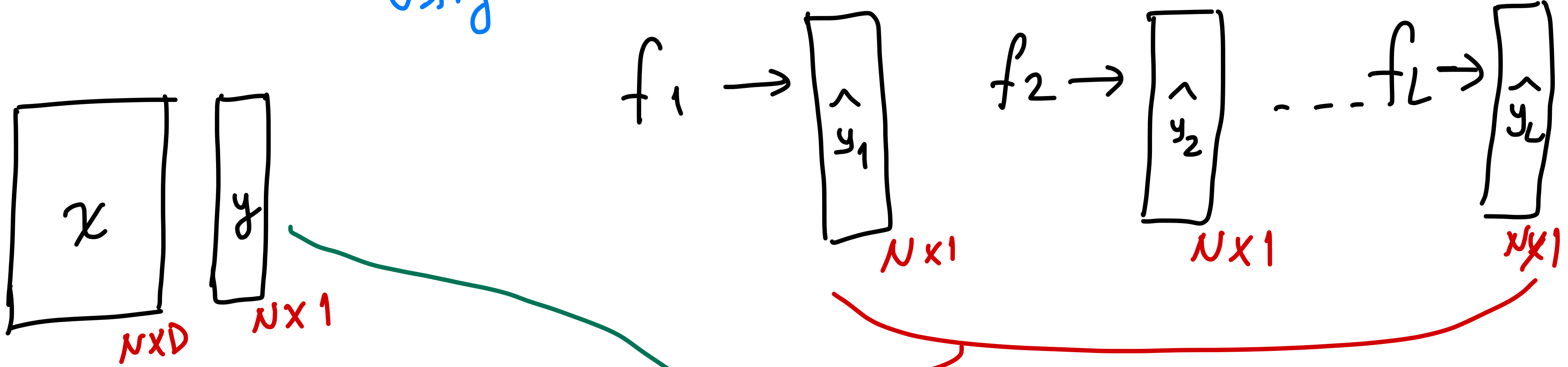
if $f(x_{N+1}) < 0 \Rightarrow -$

⇓

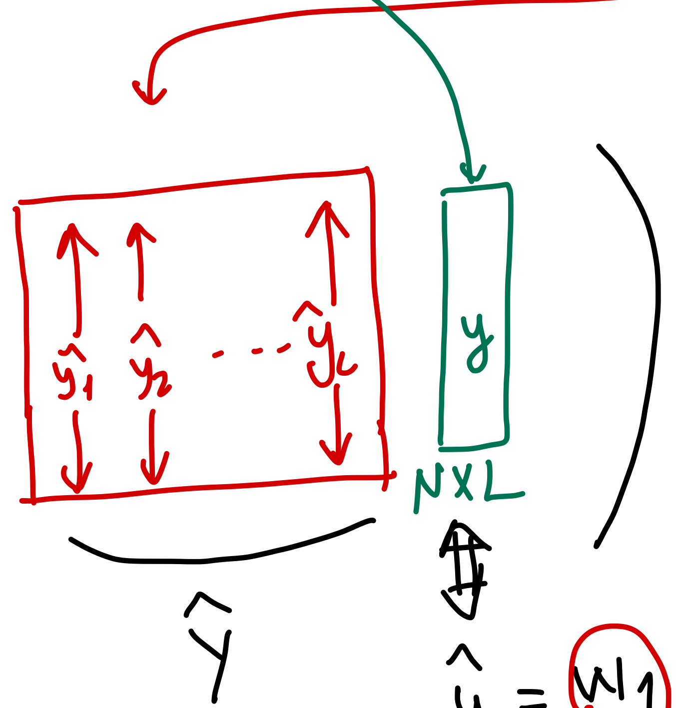
assumption:

$f_1(\cdot), f_2(\cdot), \ldots f_L(\cdot) \rightarrow$ produces either $+$ or $-$

Global Fusion : We can learn $w_1, w_2, \dots, w_L$ using another learner.



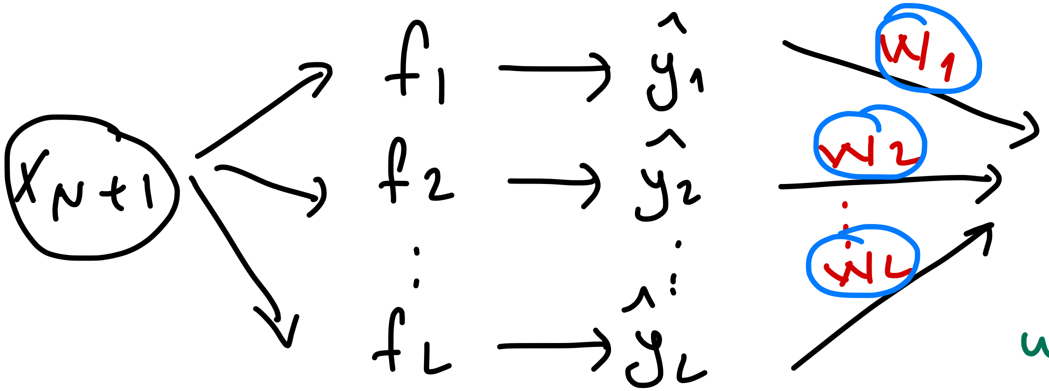
Note that $w_1, w_2, \dots, w_L$ are not functions of x_{N+1}.



can be cast into a linear regression problem

$$\hat{y} = w_1 \hat{y}_1 + w_2 \hat{y}_2 + \dots + w_L \hat{y}_L$$

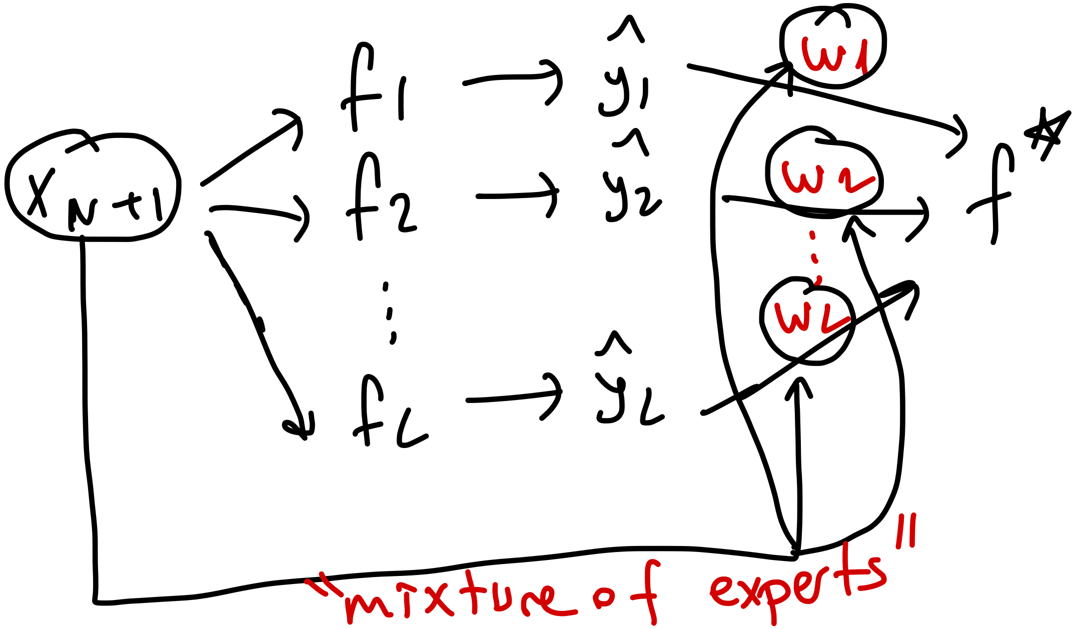
regression coefficients



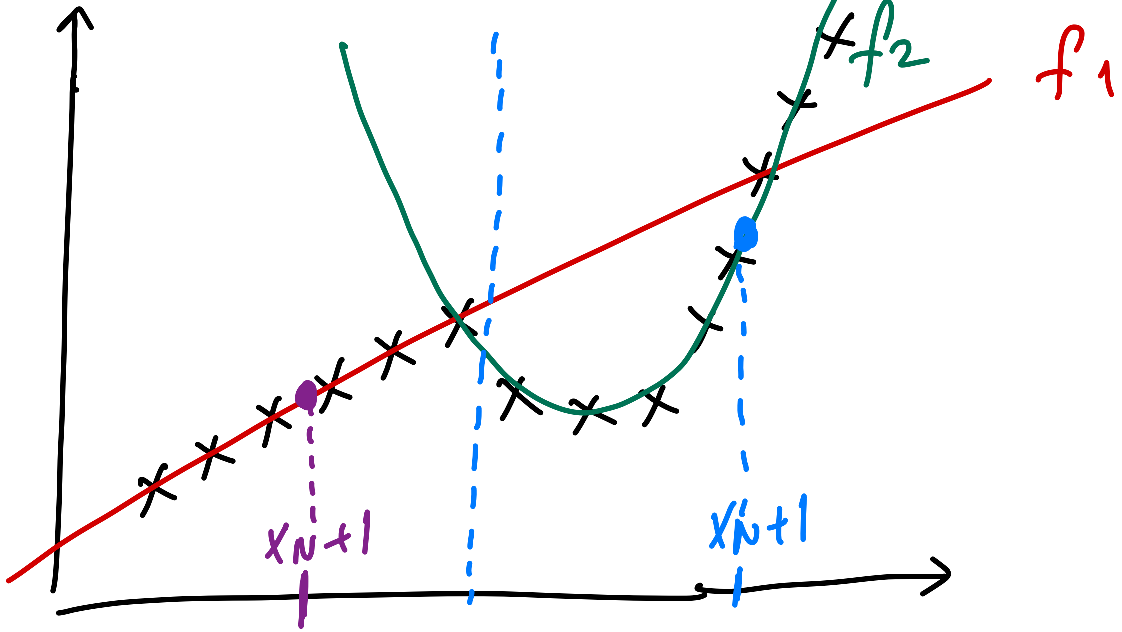
$$f^* = w_1 \hat{y}_1 + w_2 \hat{y}_2 + \dots + w_L \hat{y}_L$$

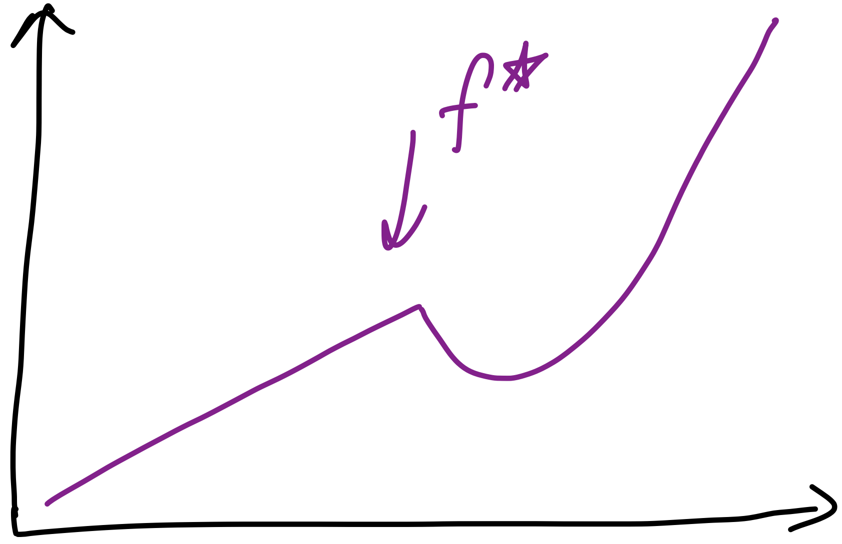
$$w_1 = w_2 = \dots = w_L = \frac{1}{L}$$

Local Fusion: $w_1, w_2, \dots w_L$ are functions of x_{N+1}.



$$f^* = w_1(x_{N+1}) \cdot \hat{y}_1 + w_2(x_{N+1}) \cdot \hat{y}_2 + \dots + w_L(x_{N+1}) \cdot \hat{y}_L$$





$w_1 \gg w_2$ $\qquad$ $w_1 \ll w_2$

$w_2 \cong 0$ $\qquad$ $w_1 \cong 0$

$$w_1(x_{N+1}) \gg w_2(x_{N+1})$$

$$w_2(x_{N+1}) \cong 0$$

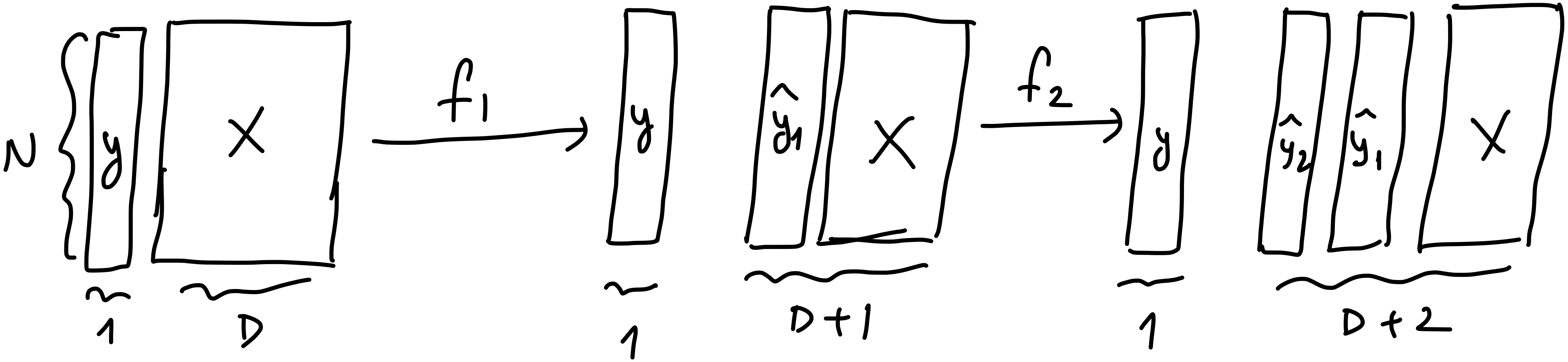
$$w_1(x_{N+1}) \ll w_2(x_{N+1})$$

$$w_1(x_{N+1}) \cong 0$$

$$w_1(x) = \frac{\exp(ax)}{\exp(ax) + \exp(bx)}$$

$$w_2(x) = \frac{\exp(bx)}{\exp(ax) + \exp(bx)}$$

MULTISTAGE COMBINATION: (serial approach)



$$x_{N+1} \to f_1(x_{N+1}) \to f_2\left(\left[f_1(x_{N+1}), x_{N+1}\right]\right) \to f_3\left(\left[f_2(x_{N+1})\ f_1(x_{N+1})\ x_{N+1}\right]\right) \cdots$$

Let us say we have L base-learners

$f_j(x)$ $\quad f_1, f_2, \ldots, f_L$

$$\hat{y} = f(f_1, f_2, \ldots, f_L \mid \Phi)$$

↳ combination function

↳ combination parameters

VOTING: $\hat{y}_i = \sum_{j=1}^{L} w_j f_j(x_i)$ } linear opinion models, ensembles

Convex combination $\Rightarrow$ $\boxed{\begin{array}{l} w_j \geq 0 \quad \forall j \\ \sum_{j=1}^{L} w_j = 1 \end{array}}$

linear combination $\Rightarrow$ $\boxed{w_j \in \mathbb{R} \quad \forall j}$