COMBINING MULTIPLE LEARNERS

- many different algorithms/learners
- NO FREE LUNCH THEOREM ⇒ no single algorithm is always the best one

- several algorithms → k-NN ($k=3$, $k=5$, $k=7$, ...)
- several hyperparameters → MLP ($H=10$, $H=20$, $H=50$)

- 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	−	−	...	−

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 → MLP + k-NN
"inductive bias" → MLP + k-NN + SVM + DT
→ one parametric + one nonparametric

② Different Hyperparameters

k-NN → k=3 (local), k=17 (global)

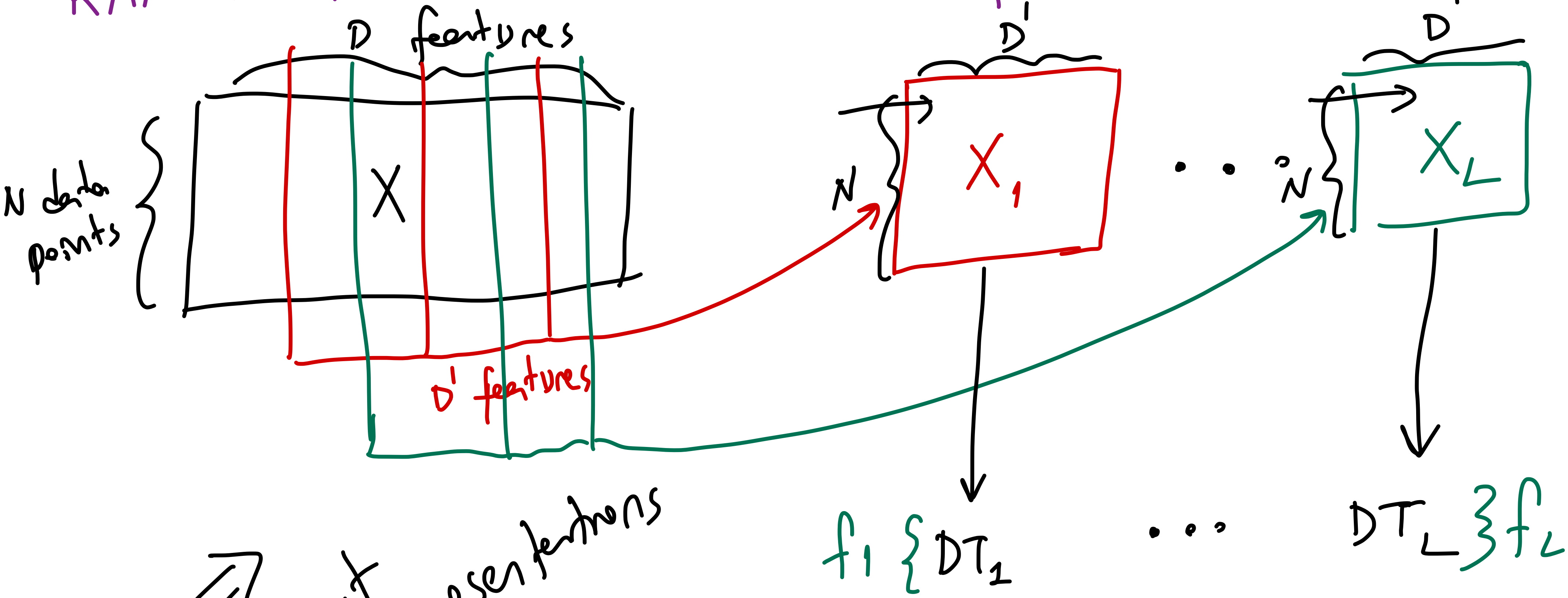
MLP → H=50 (simpler), H=500 (more complex)

③ Different Input Representations "multiview" or "multimodal"

different types of sensors/measurements/modalities/representations

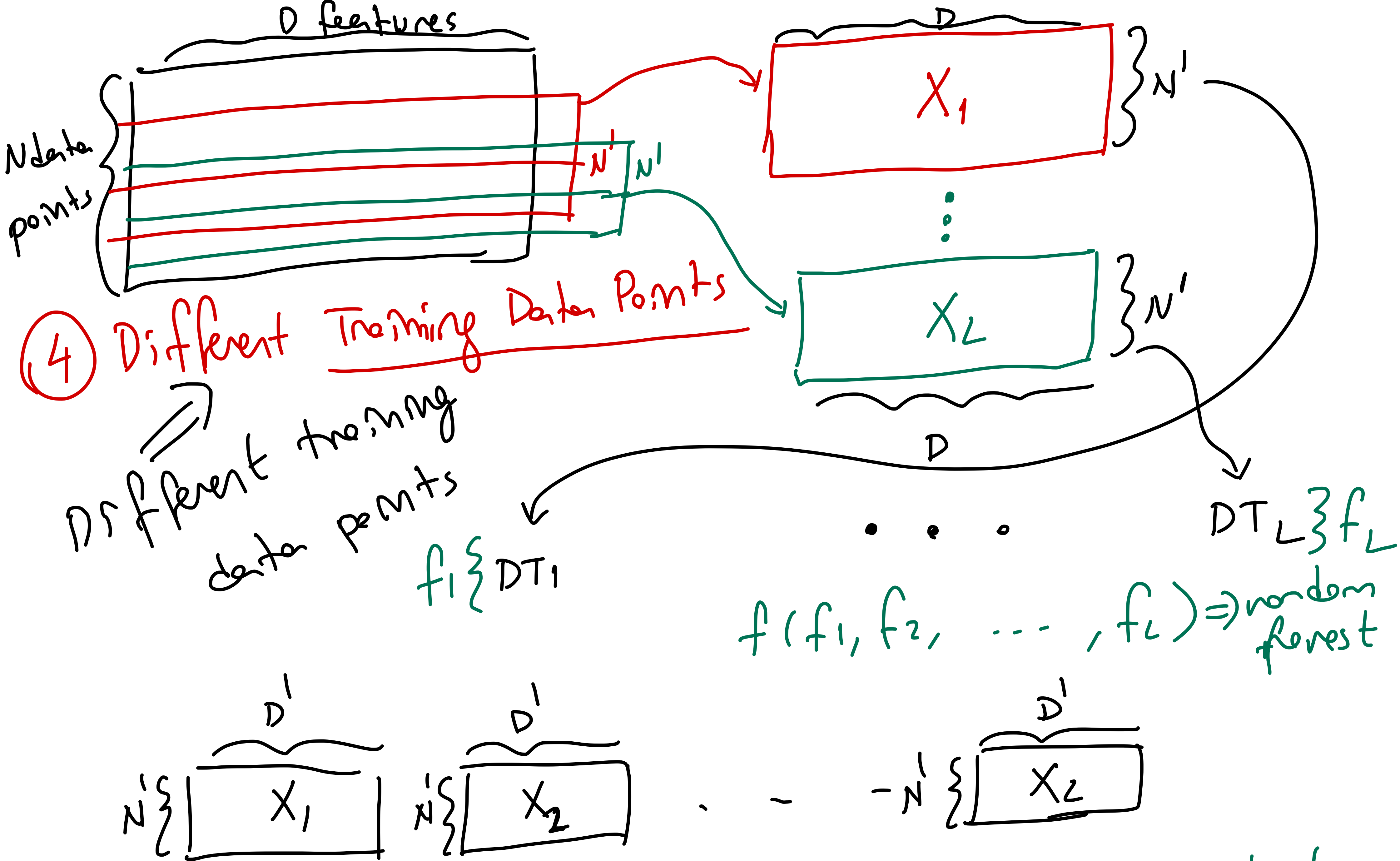
sensor fusion ⇒ audio + video

RANDOM FOREST (a collection of decision trees)



⇗ Different input representations

$f(f_1, f_2, \ldots, f_L) \Rightarrow$ random forest



randomly select subsets of rows and columns

Model Combination Strategies:

multiple [expert / learner / base-learner / algorithm] combination

$L = \#$ of base learners

global combination (learner fusion)

local combination (learner selection)

$f_1 \quad f_2 \quad \ldots \quad f_L$

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

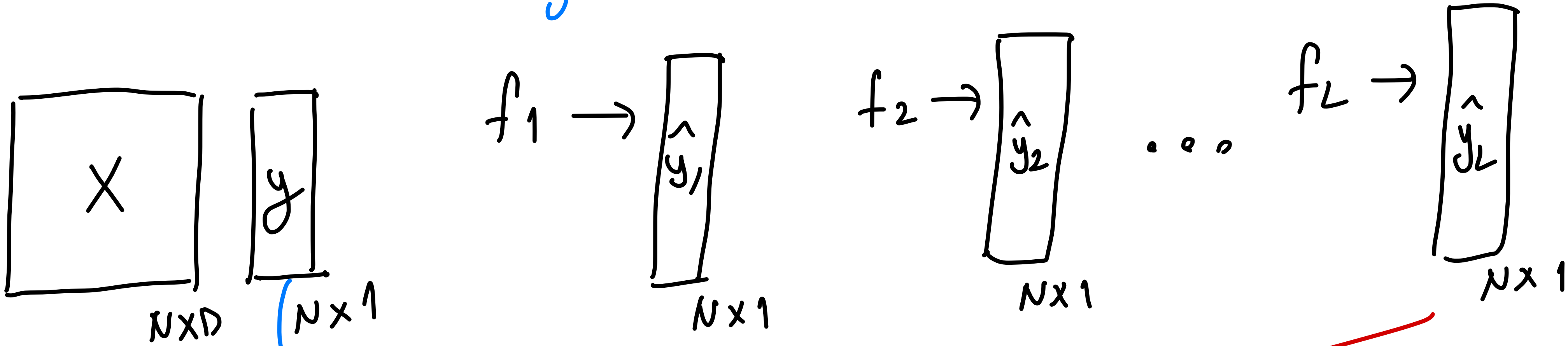
$f_1(x_{N+1}) \quad f_2(x_{N+1}) \quad \ldots \quad f_L(x_{N+1})$

Combination $\Rightarrow$ $w_1 \underbrace{f_1(x_{N+1})}_{} + w_2 f_2(x_{N+1}) + \ldots + w_L f_L(x_{N+1})$

Majority Voting $\Rightarrow$ $w_1 = 1 \quad w_2 = 1 \quad \ldots \quad w_L = 1$

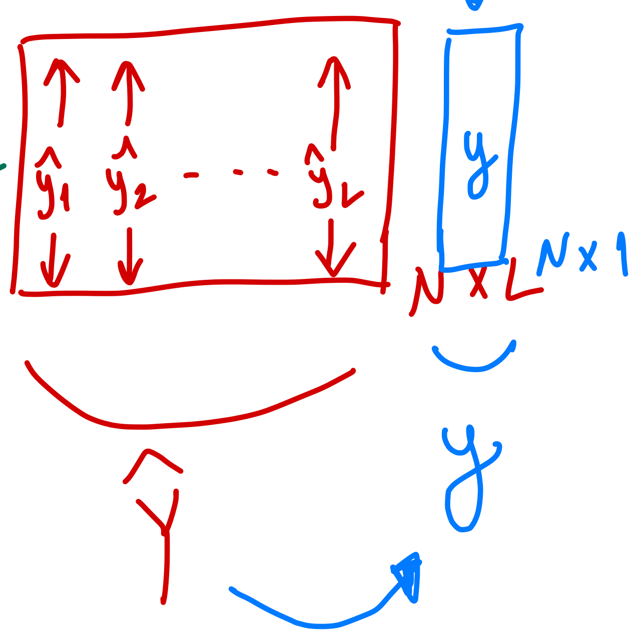
either + or − (+1 or −1)

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



predictions of base-learners

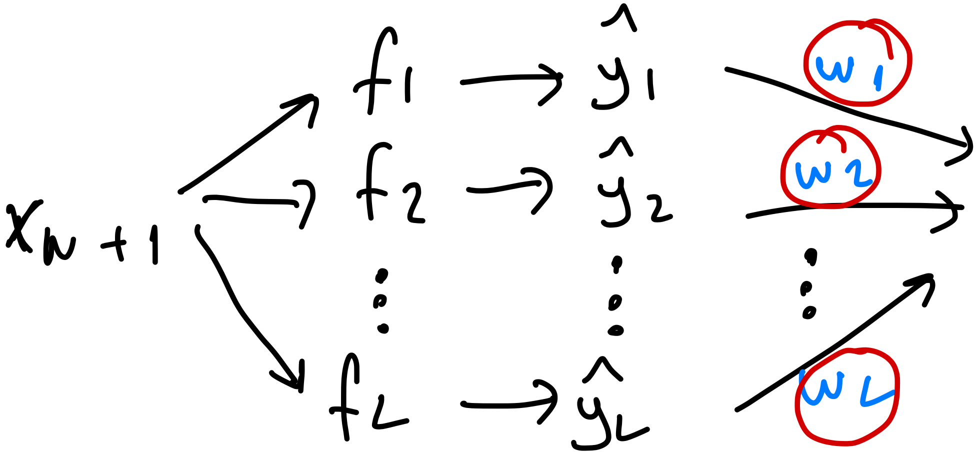
Note that $w_1, w_2, \dots, w_L$ are not functions of x_{N+1}.



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

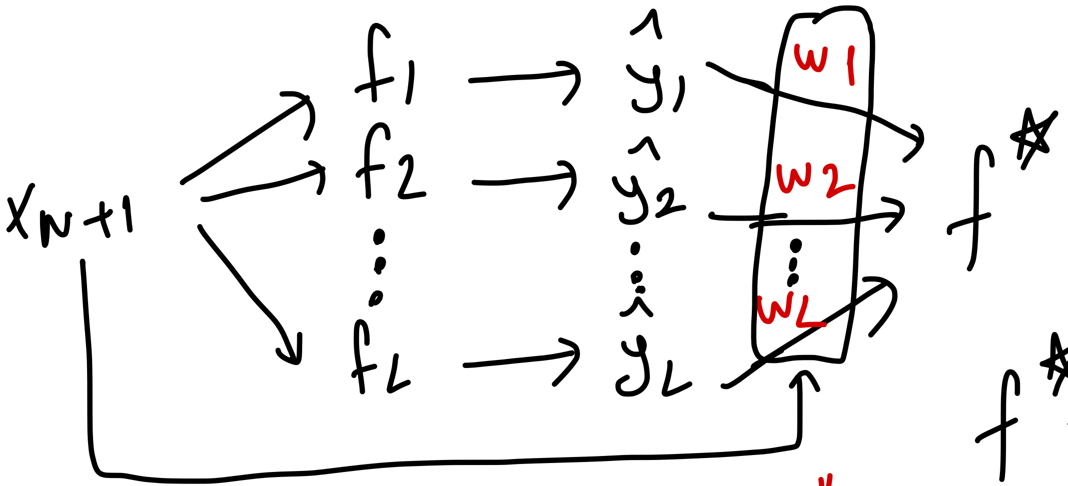


$$f^* = w_1\hat{y}_1 + w_2\hat{y}_2 + \dots + w_L\hat{y}_L$$

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

majority voting

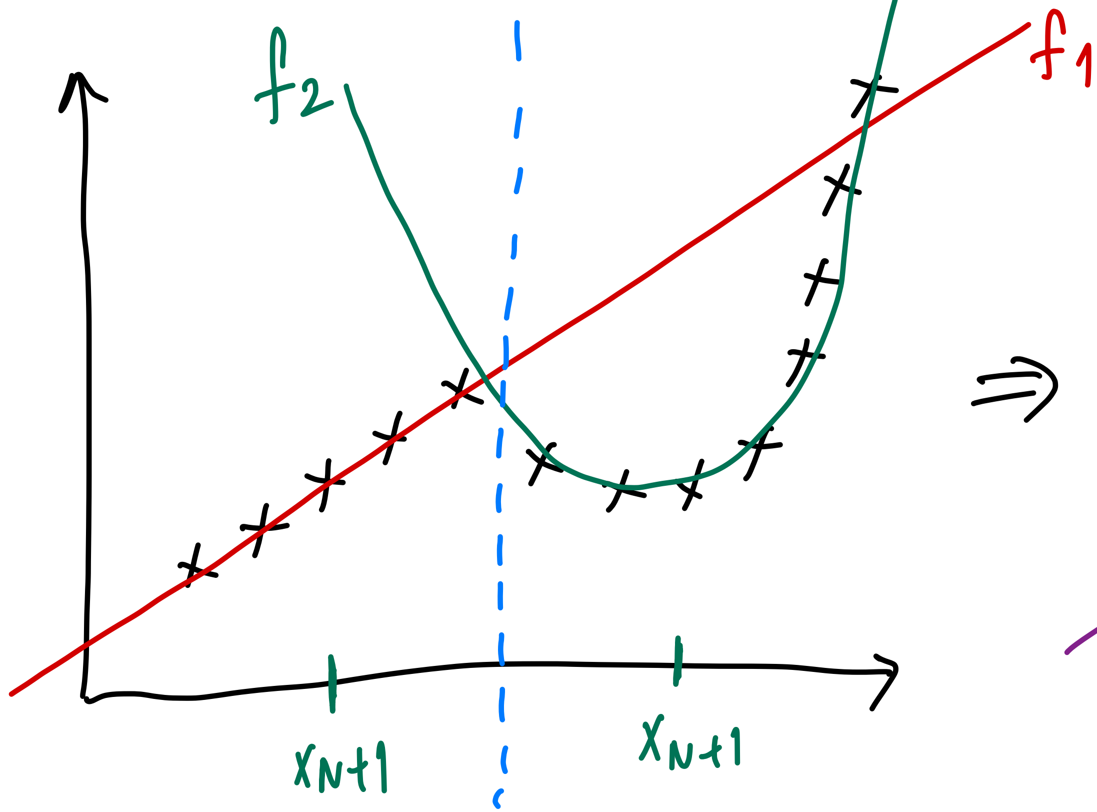
Local Fusion: $w_1, w_2, \dots, w_L$ are now functions of x_{N+1}.



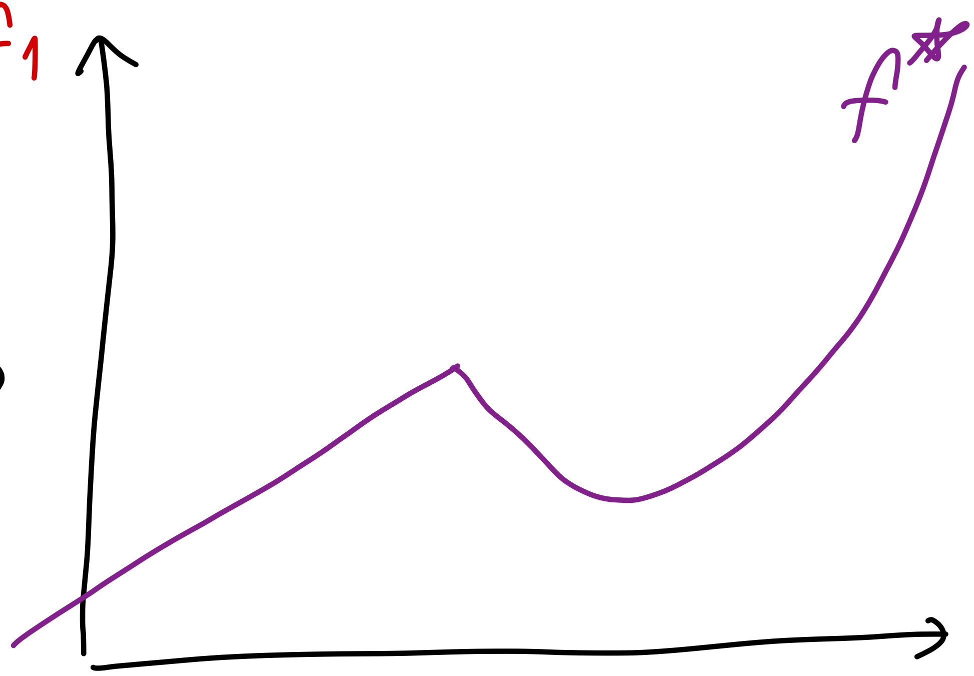
"mixture of experts"

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

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



$\Longrightarrow$



$w_1 \gg w_2$

$w_2 \cong 0$

$w_1 \ll w_2$

$w_1 \cong 0$

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

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

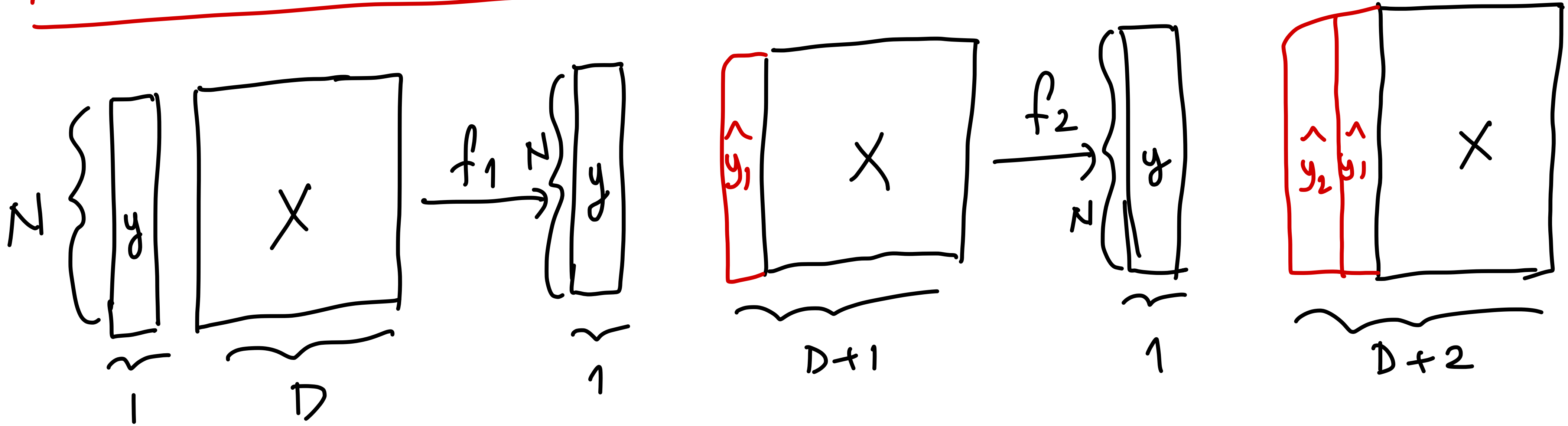
$$w_1(x_{N+1}) \ll w_2(x_{N+1})$$

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

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

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

MULTISTAGE COMBINATION (serial approach)



$$x_{N+1} \to f_1(x_{N+1}) \to f_2([f_1(x_{N+1}), x_{N+1}]) \longrightarrow \ldots$$

Let us say we have L base learners

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

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

↳ combination function

↳ combination parameters

VOTING:

$$\hat{y}_i = \sum_{j=1}^{L} w_j f_j(x_i)$$

linear opinion models, ensembles.

Convex combination $\Rightarrow$ $w_j \geq 0 \quad \forall j$

$$\sum_{j=1}^{L} w_j = 1$$

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