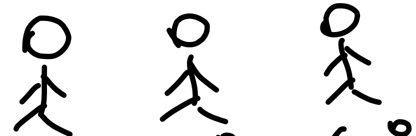
# Combining Multiple Learners



60% 60% 60%

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

$$Pr(-++)+Pr(+-+)+Pr(++-)+Pr(+++)$$

$$3\cdot\frac{6}{10}\frac{6}{10}\frac{4}{10}+\frac{6}{10}\frac{6}{10}\frac{6}{10}$$

$$=\frac{18.36}{1000}=\frac{648}{1000}\cong 65\%$$

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

- MAIN IDEA ⇒ DIVERSITY

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

If they produce $\frac{\text{the same}}{\text{very similar}}$ 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 negative 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)

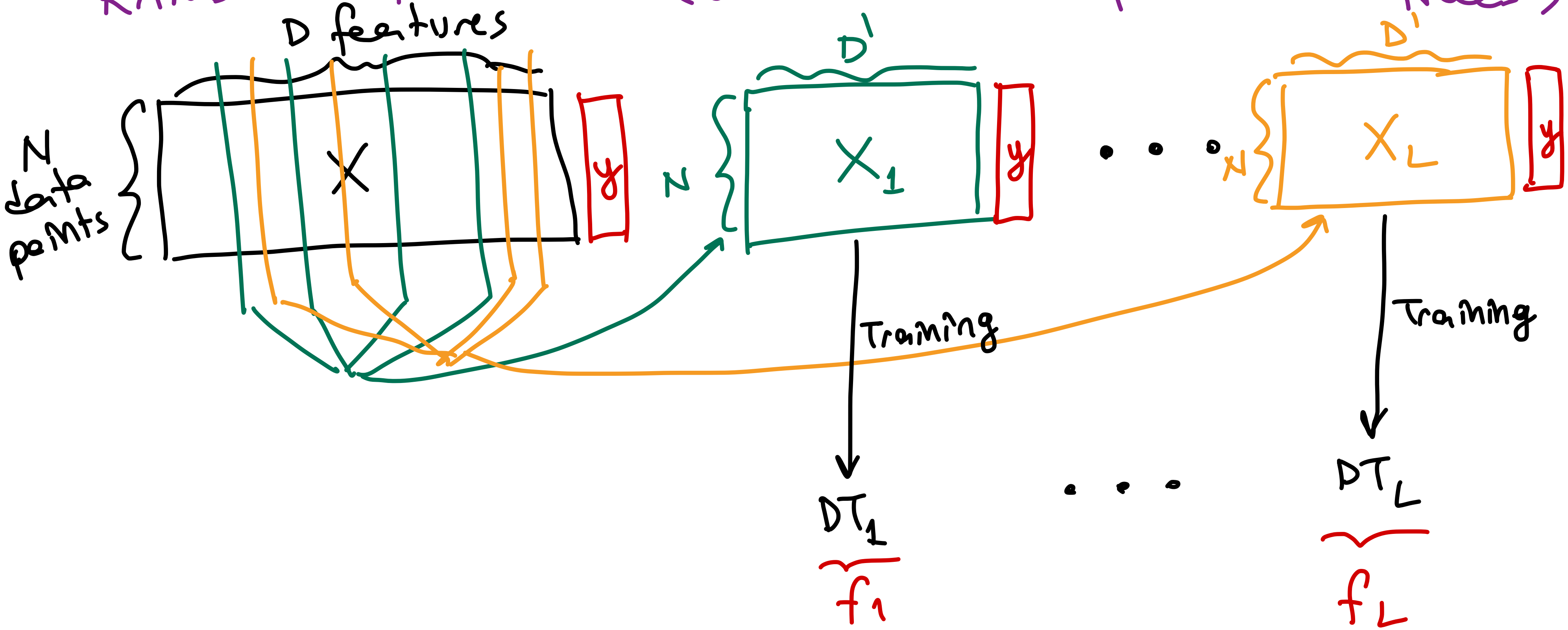
MLP
- H=50 (simpler)
- H=500 (more complex)

③ Different input representations (views, modalities, measurements, sensors)

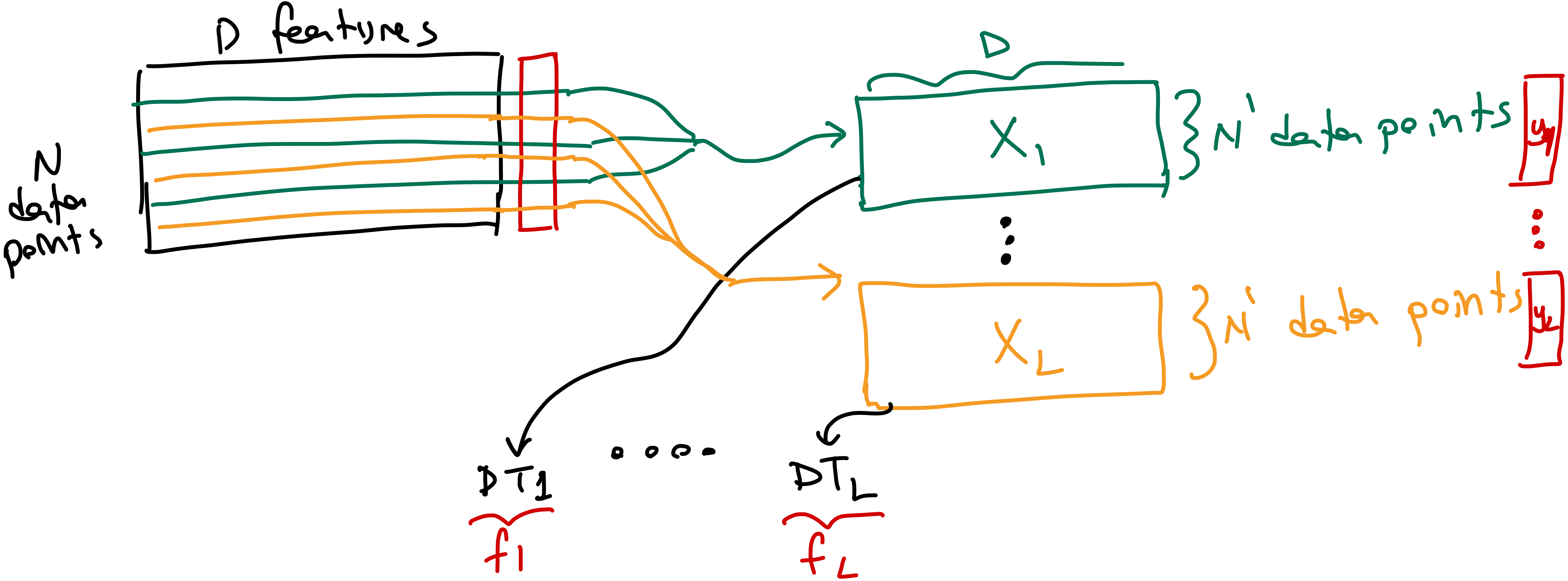
sensor fusion ⇒ audio + video


trained A/ML Model = f (Algorithm ✓, Hyperparameters ✓, Training Data ?)

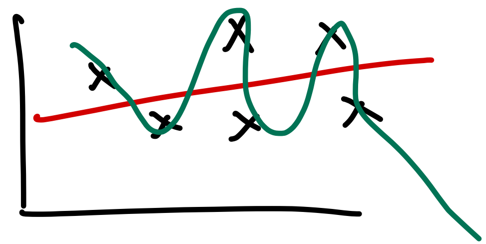
# RANDOM FORESTS (a collection of decision trees)



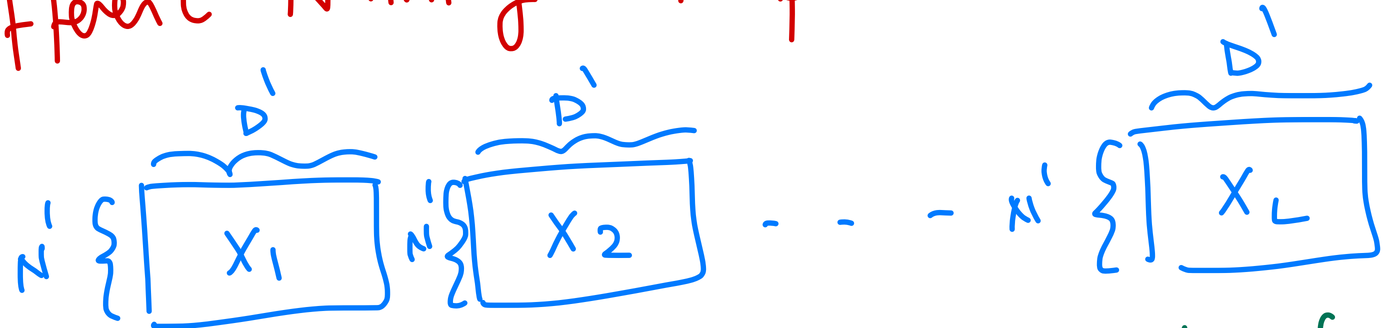

$$f(f_1, f_2, \ldots, f_L) \Rightarrow \text{random forest}$$




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



④ Different training data points




randomly select subsets of rows and columns

# Model Combination Strategies:

multiple [expert / learner / base-learner / algorithm] combination → global combination (learner fusion)

→ local combination (learner selection)

$L = \#$ of base-learners

models → $f_1 \quad f_2 \quad \ldots\ldots \quad f_L$

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

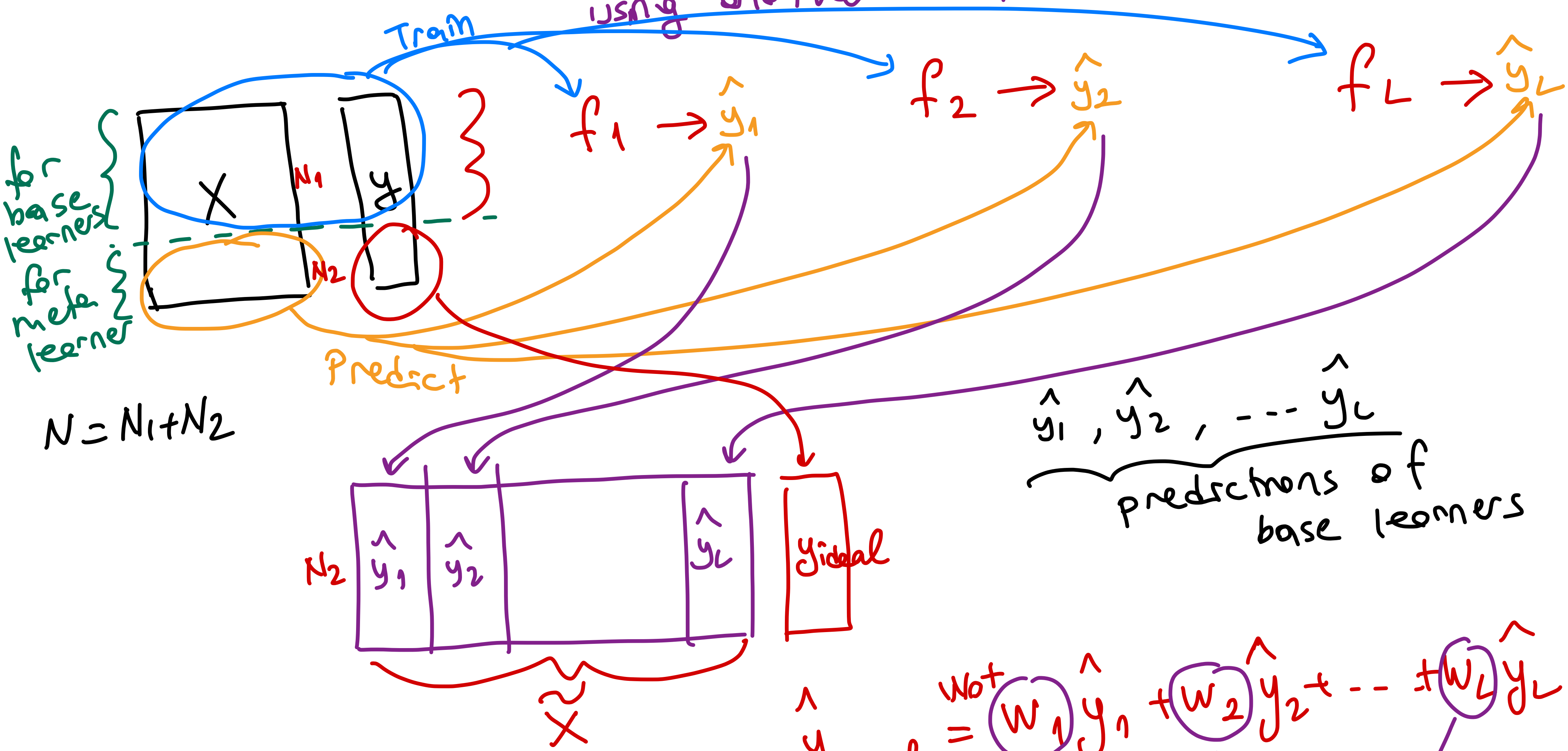
predictions → $f_1(x_{N+1}) \quad f_2(x_{N+1}) \quad \ldots\ldots \quad f_L(x_{N+1})$

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

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

learners predict $+1$ or $-1$.

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

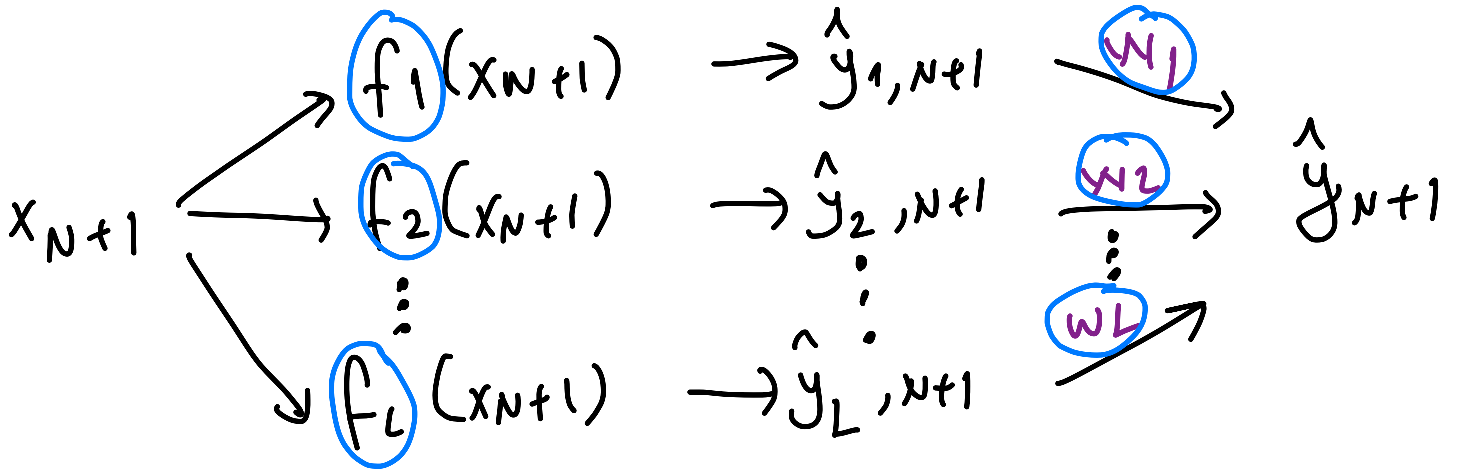



$$\hat{y}_{ideal} = w_0 + w_1 \hat{y}_1 + w_2 \hat{y}_2 + \dots + w_L \hat{y}_L$$

regression coefficients

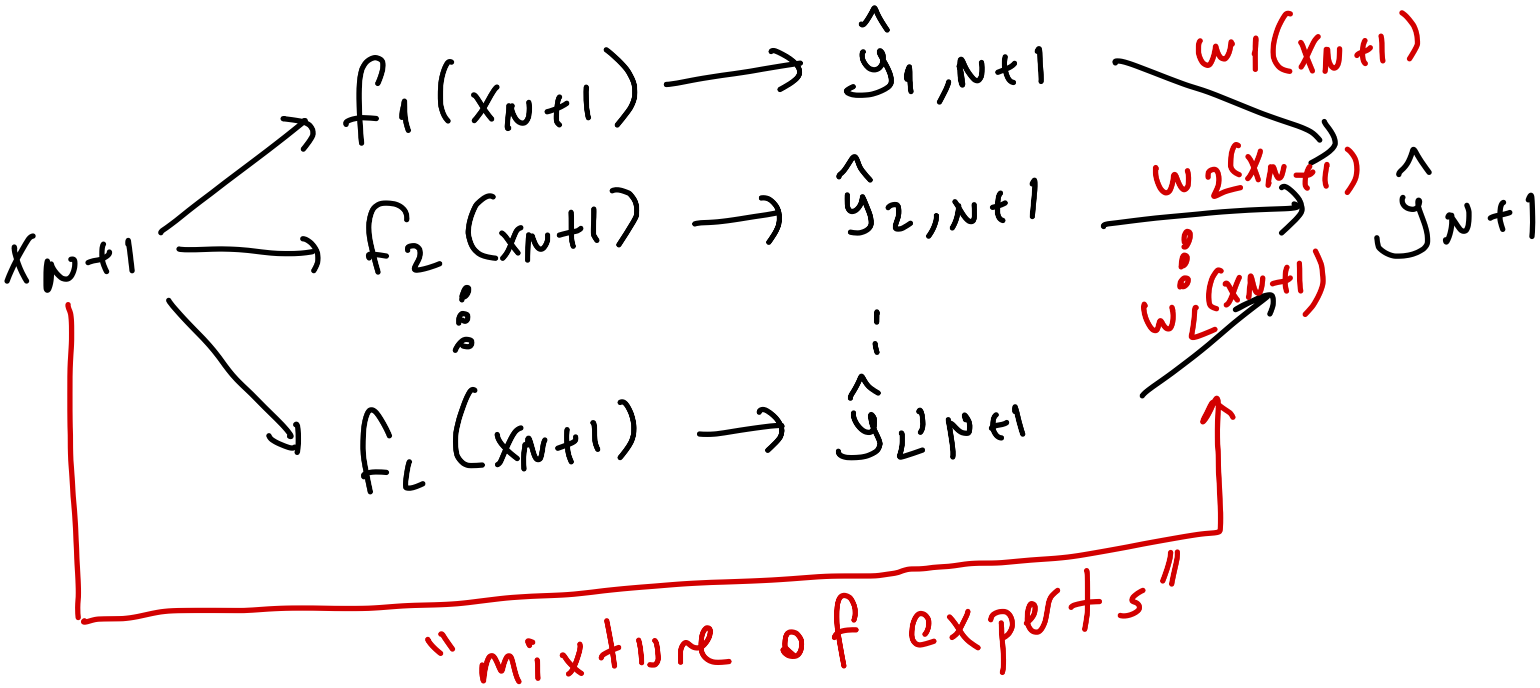
$\tilde{X} \rightarrow y_{ideal}$

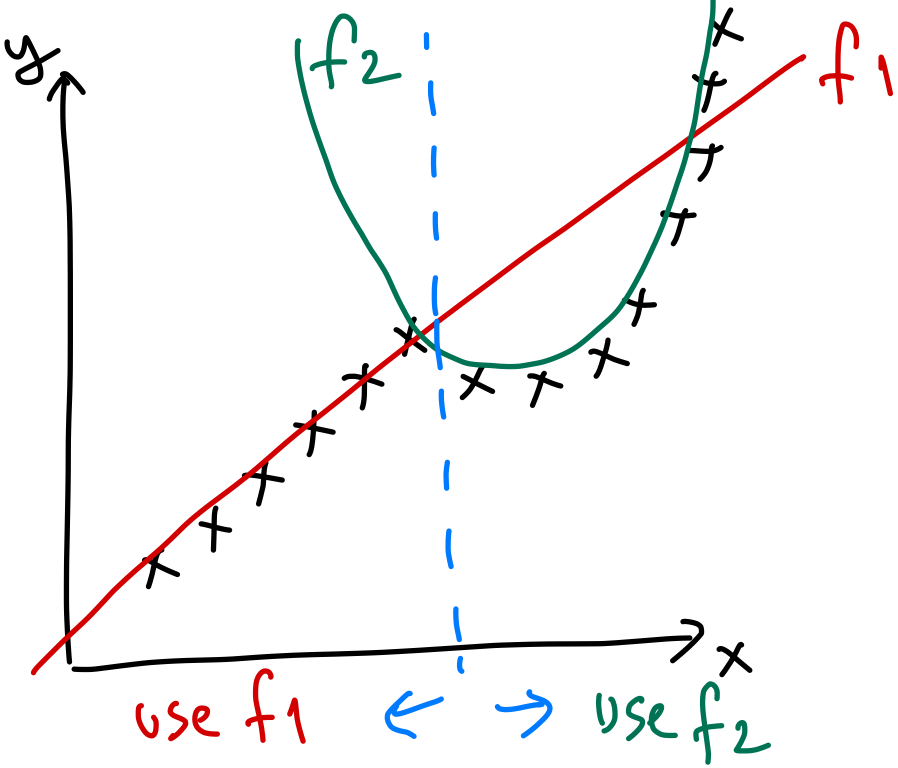
a new supervised learning problem.

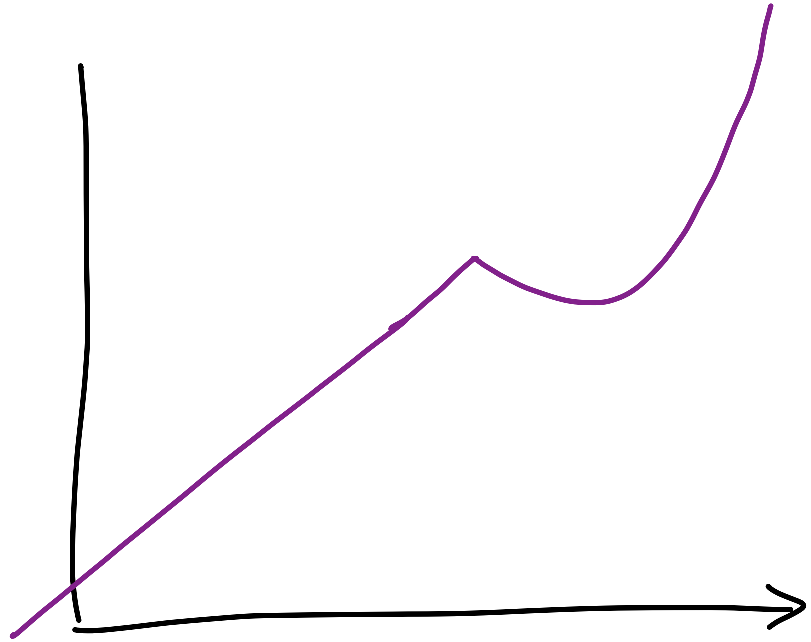



$w_1, w_2, \ldots, w_L$ are not functions of $x_{N+1}$.

Local fusion: $w_1, w_2, \ldots w_L$ are new functions of $x_{N+1}$.





$w_1 = 1$ $\quad$ $w_1 \simeq 0$

$w_2 \simeq 0$ $\quad$ $w_2 = 1$

$$w_1(x) + w_2(x) = 1$$

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

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

learn a & b on a separate dataset.