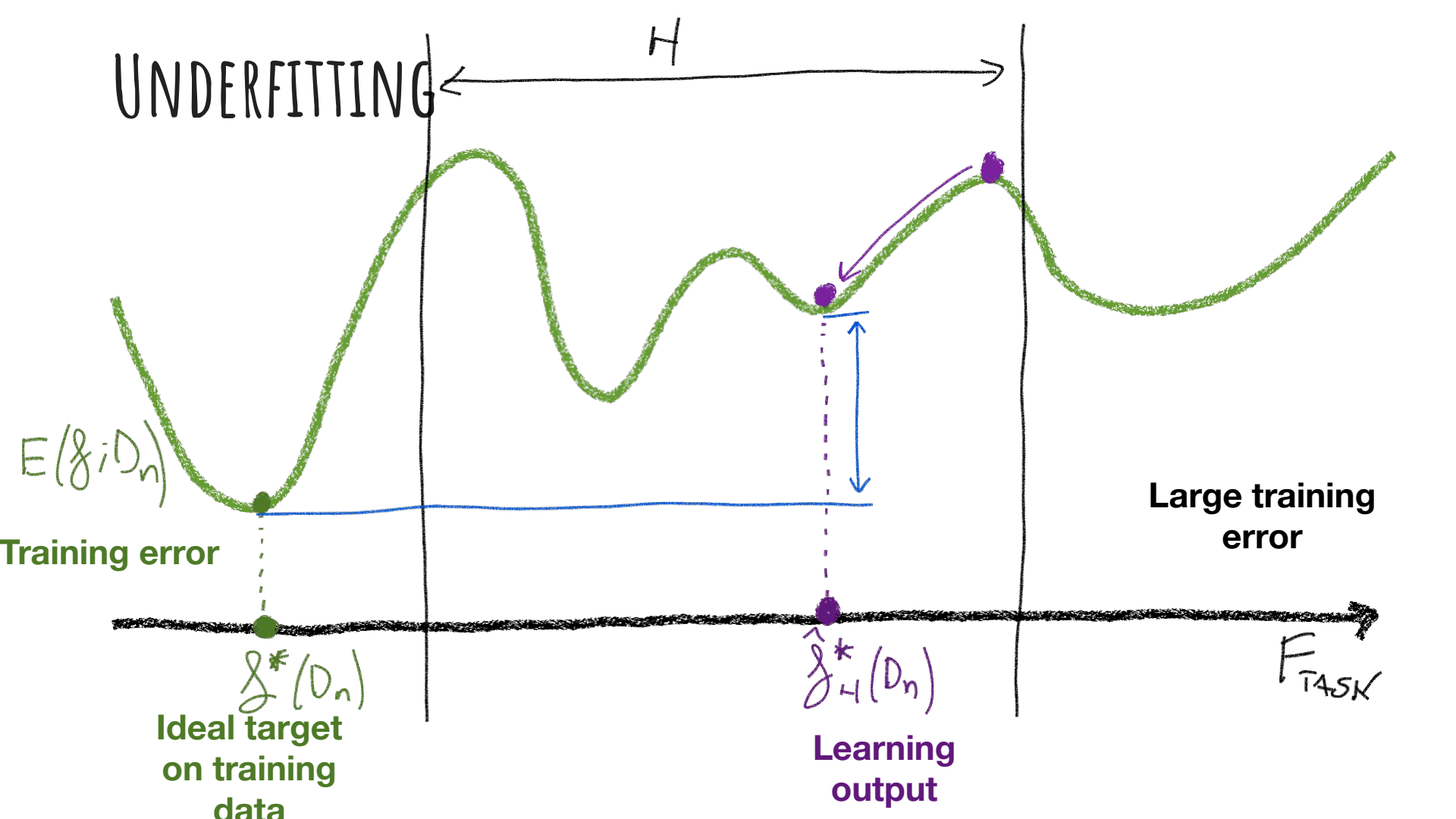
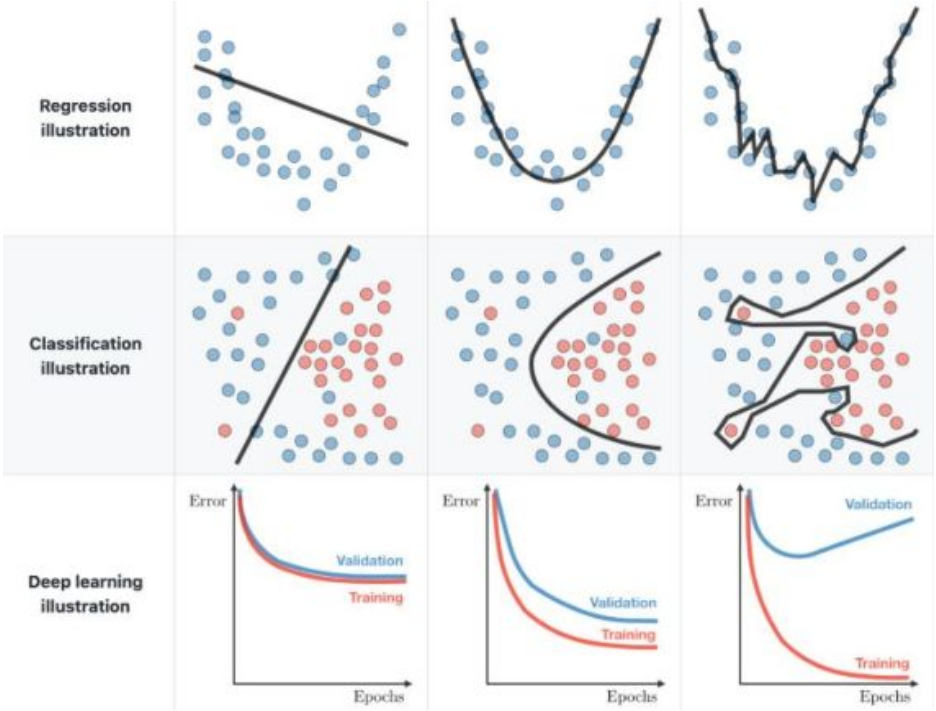

UNDERFITTING
H
$E(g; D_n)$
Training error
Large training error
$g^*(D_n)$
Ideal target on training data
$\hat{g}^*_H(D_n)$
Learning output
F_{TASK}

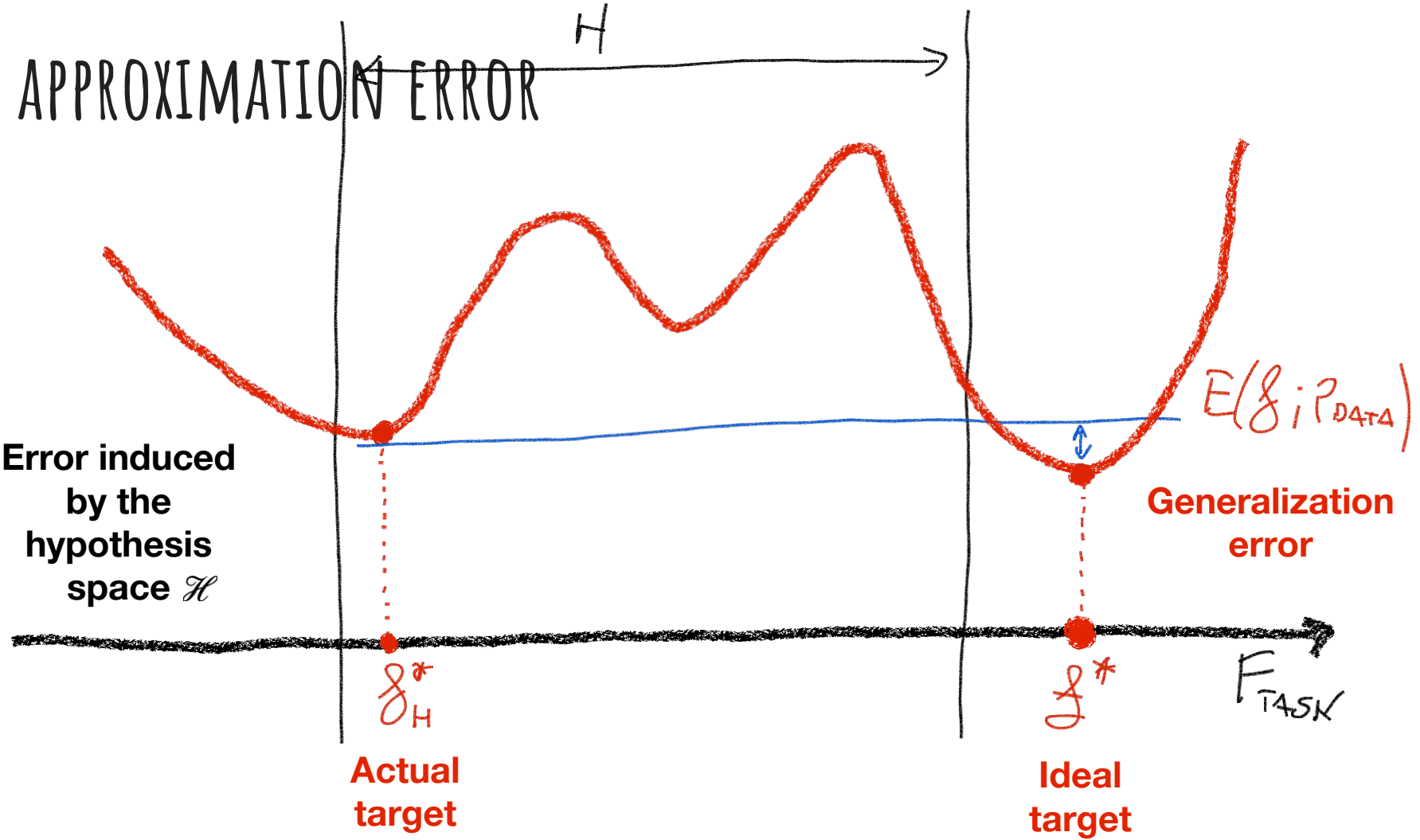
Overfitting & Underfitting

COMMON QUESTION
explain the concept of overfitting and underfitting with examples
- regression
- classification

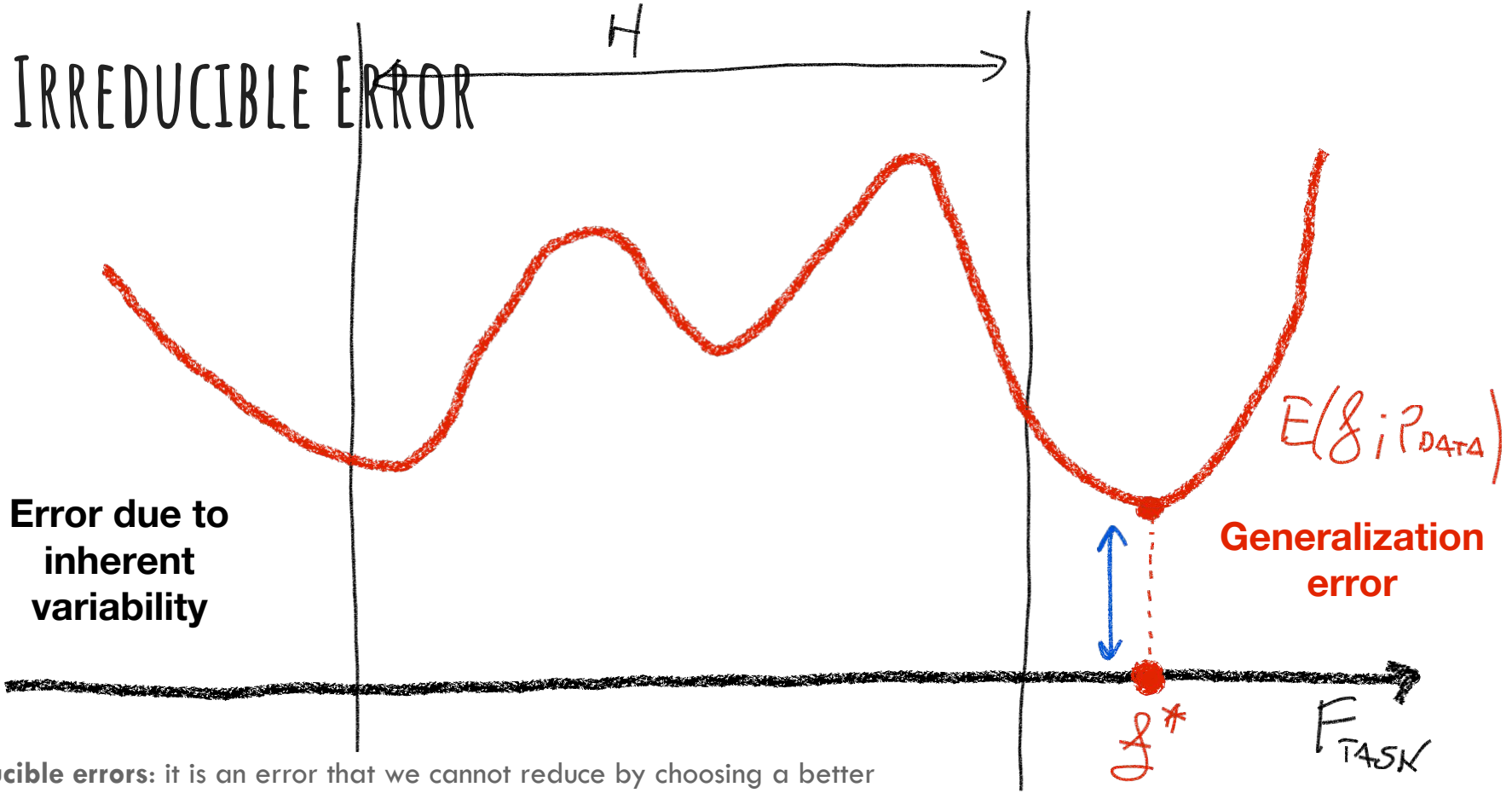
UNDERFITTING CORRECT OVERFITTING

Example (e.g. neural network training process)



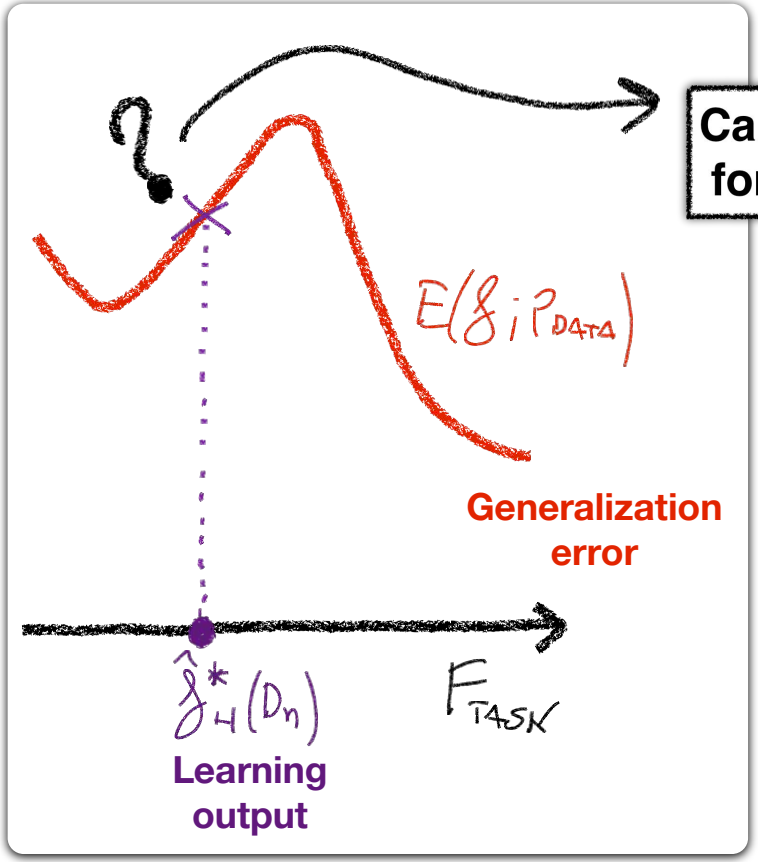

APPROXIMATION ERROR
H
Error induced by the hypothesis space $\mathcal{H}$
$E(g; P_{DATA})$
Generalization error
g^*_H
g^*
F_{TASK}
Actual target
Ideal target

Irreducible Error



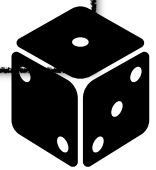
Irreducible errors: it is an error that we cannot reduce by choosing a better model. it is due to randomness or natural variability in a system.
These are deterministic factors that you don't observe, but which could be interpreted as inherent variability.

HOW TO ESTIMATE GENERALIZATION ERROR



Cannot be computed for p_{data} is unknown

We need a data sample



TRAINING SET

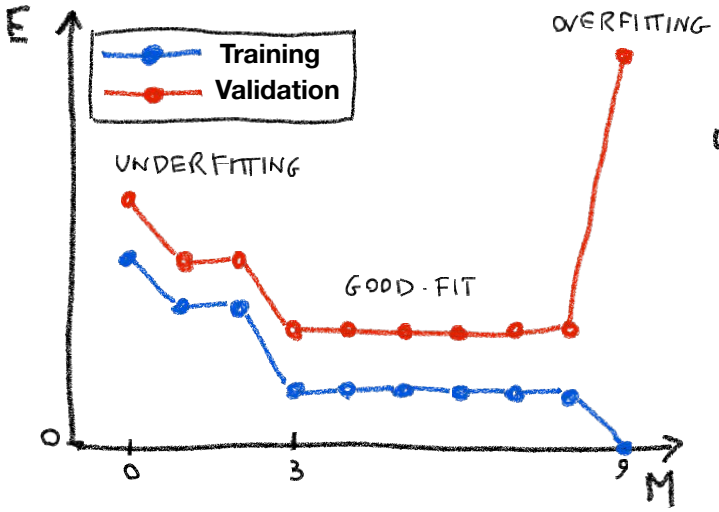
VALIDATION SET

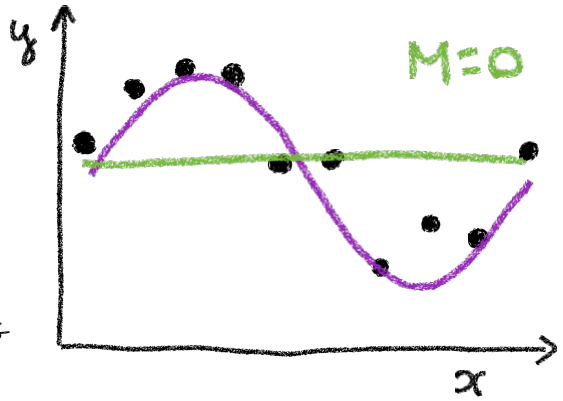
TEST SET

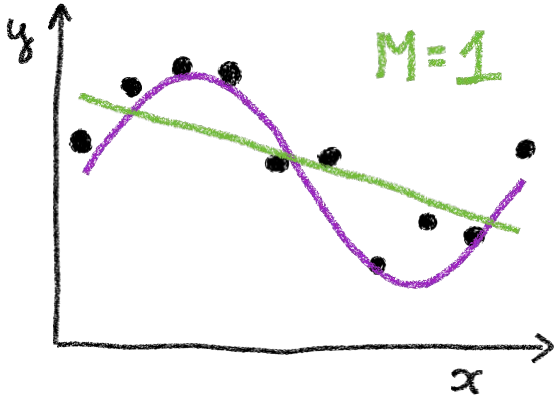
Used in the process of selecting the model (biased)

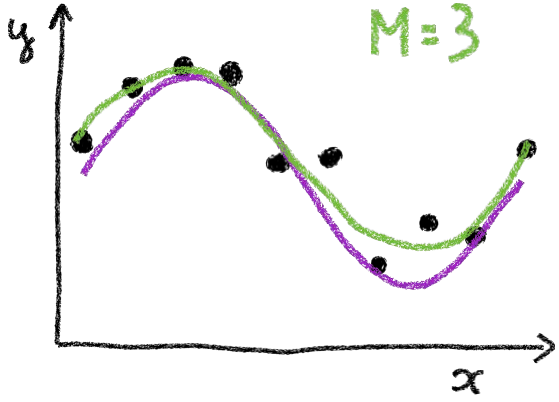
Example: polynomial curve fitting

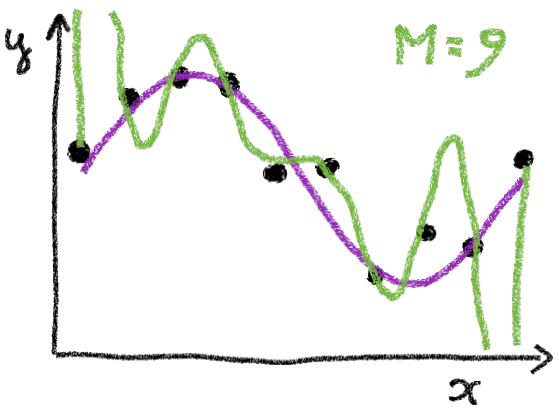
Solutions for different model complexities











How to improve generalization

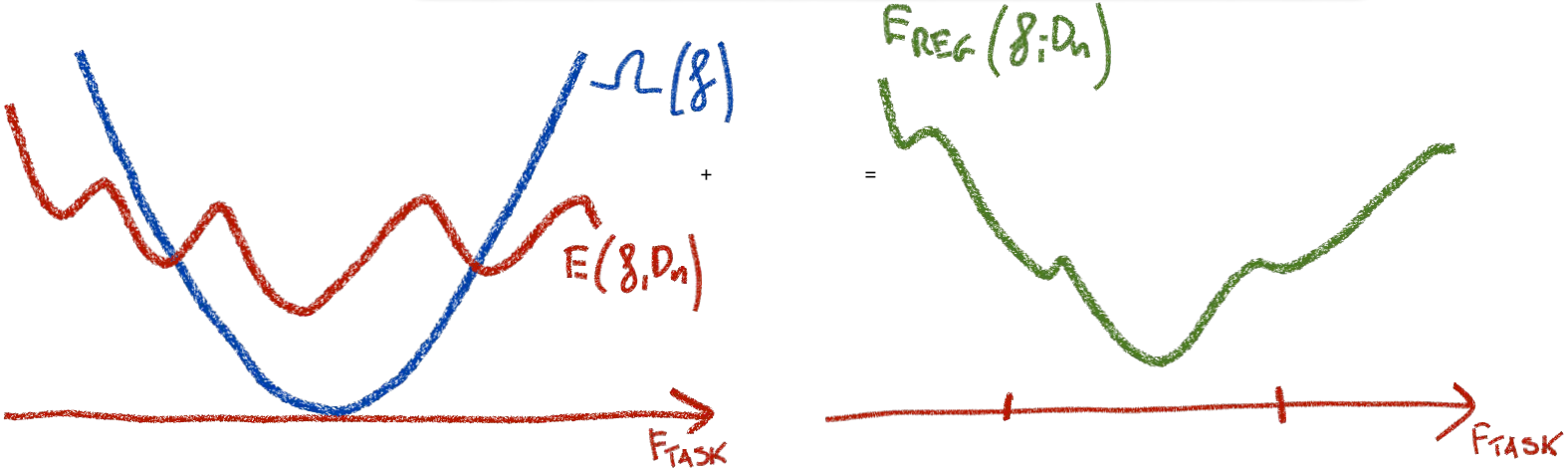
- Avoid attaining the minimum on training error
- Reduce model capacity
- Change the objective with a regularization term
- Inject noise in the learning algorithm
- Stop the learning algorithm before convergence

Regularization

Modification of the training error function with a term $\Omega(f)$ that typically penalizes complex solutions

TRADE-OFF PARAMETER

$$E_{\text{reg}}(f; \mathcal{D}_n) = E(f; \mathcal{D}_n) + \lambda_n \Omega(f)$$



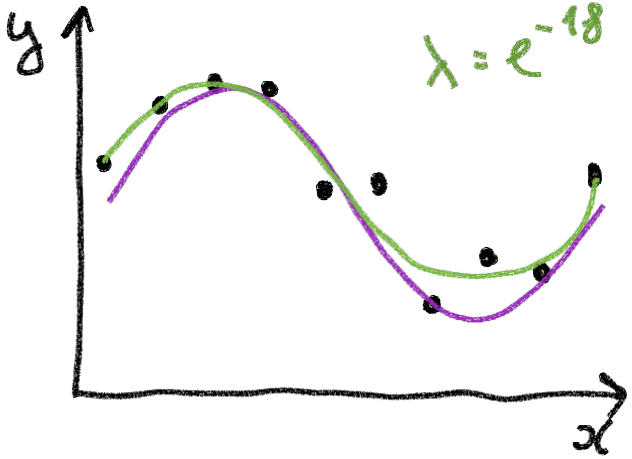
Example: polynomial curve fitting

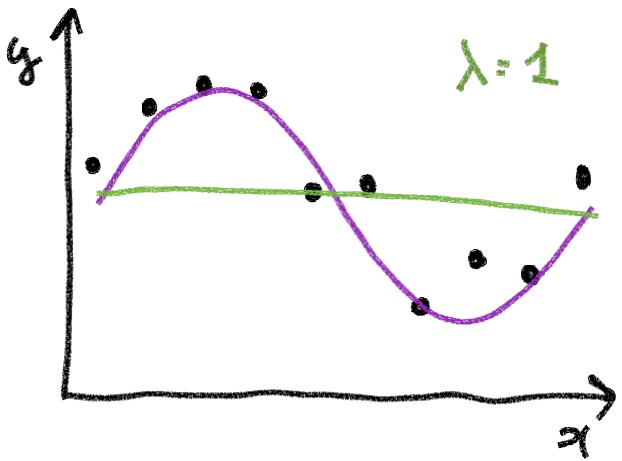
We regularize by penalizing polynomials with large coefficients

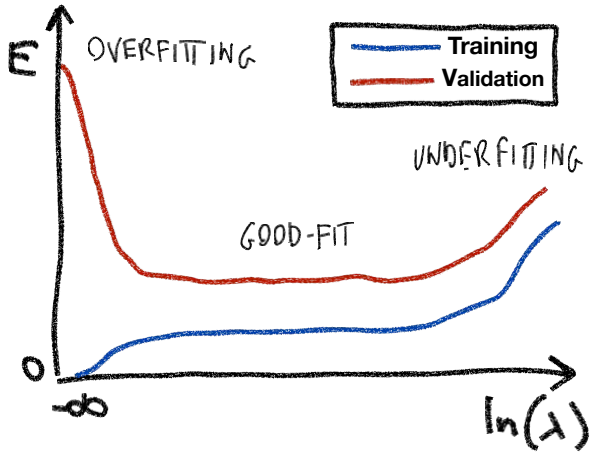
esistono altri tipi di normalizer

$$E_{\text{reg}}(f_w ; \mathcal{D}_n) = \frac{1}{n} \sum_{i=1}^{n} [f_w(x_i) - y_i]^2 + \frac{\lambda}{n} \|w\|^2$$

$\lambda = \sum_i w_i^2$







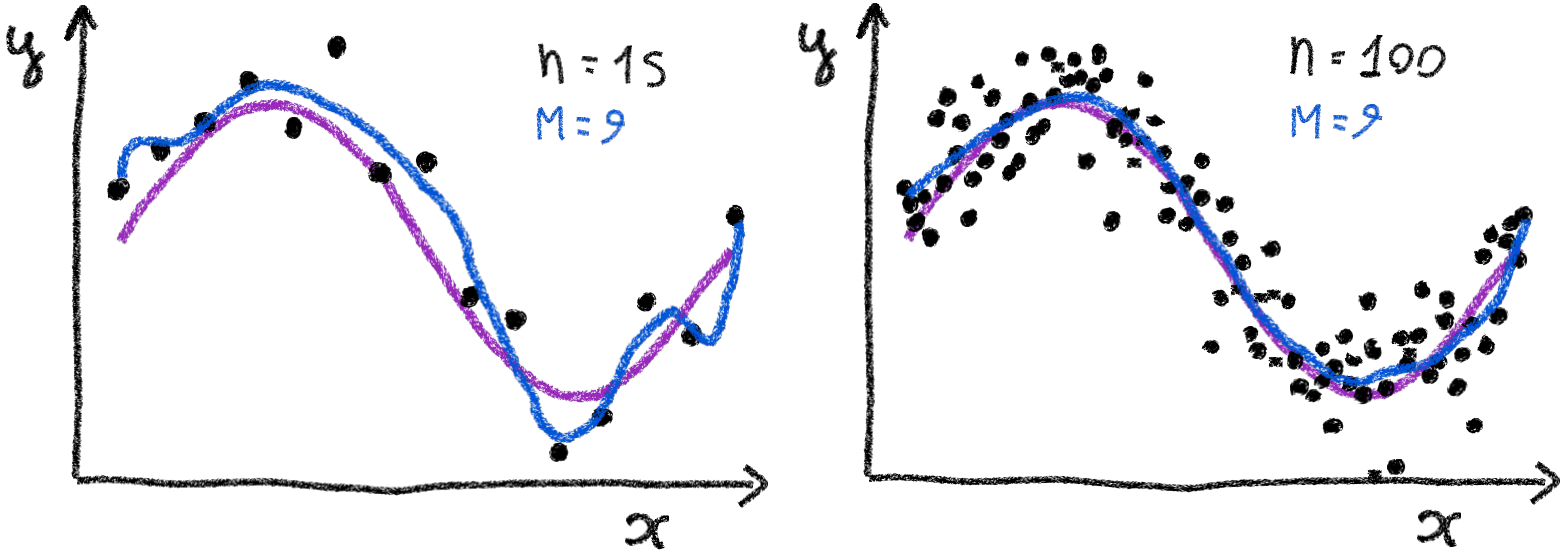
How to improve generalization

- Increase the amount of data
- Adding more training samples
- Augmenting the training set with transformations
- Combine predictions from multiple, decorrelated models (ensembling)

Example: polynomial curve fitting

Generalization vs data size

More data means avoiding overfitting



$$E(f;\mathscr{D}_n) \to E(f;p_{\text{data}}) \quad \textbf{as} \quad n \to \infty$$