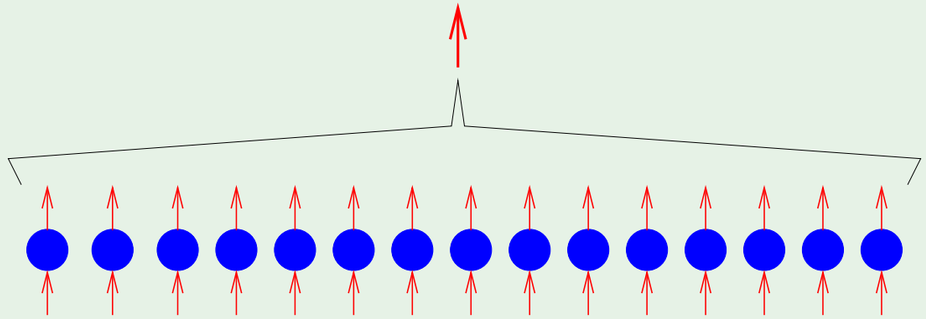
Outline

- The map of machine learning
- Bayesian learning
- Aggregation methods
- Acknowledgments

What is aggregation?

Combining different solutions $h_1, h_2, \cdots, h_T$ that were trained on $\mathcal{D}$:



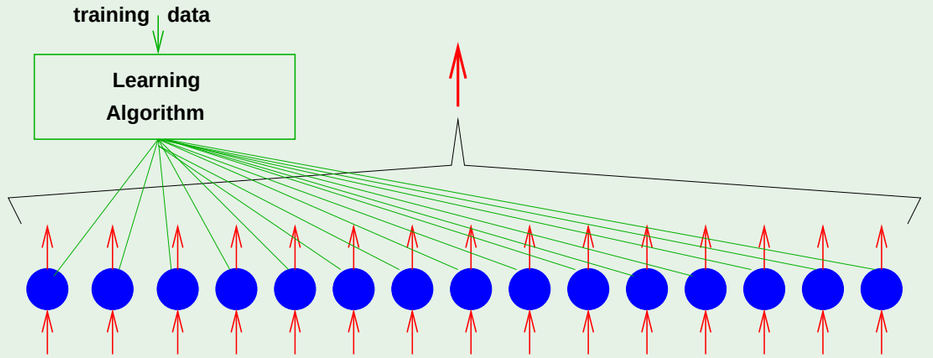
Regression: take an average

Classification: take a vote

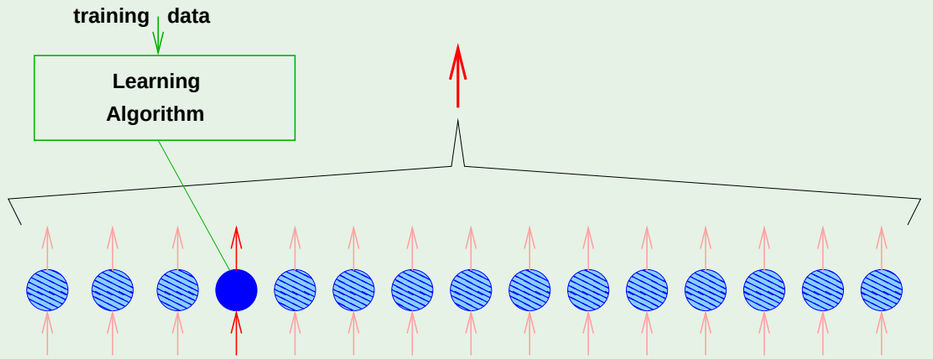
a.k.a. *ensemble learning* and *boosting*

Different from 2-layer learning

In a 2-layer model, all units learn **jointly**:



In aggregation, they learn **independently** then get combined:



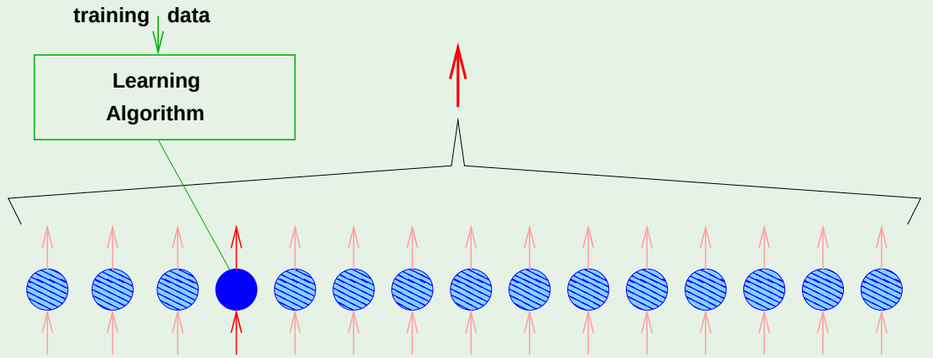
Two types of aggregation

1. **After the fact:** combines existing solutions

Example. Netflix teams merging "blending"

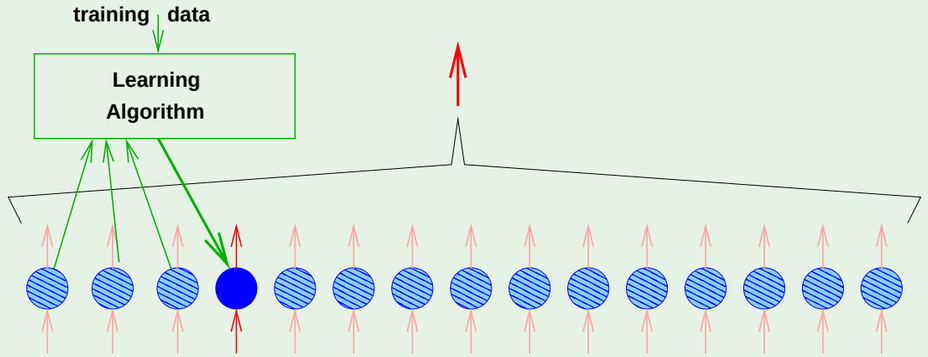
2. **Before the fact:** creates solutions to be combined

Example. Bagging - resampling $\mathcal{D}$



Decorrelation - boosting

Create $h_1, \cdots, h_t, \cdots$ sequentially: Make h_t decorrelated with previous h's:



Emphasize points in $\mathcal{D}$ that were misclassified

Choose weight of h_t based on $E_{\text{in}}(h_t)$

Blending - after the fact

For regression, $h_1, h_2, \cdots, h_T \longrightarrow g(\mathbf{x}) = \sum_{t=1}^{T} \alpha_t \, h_t(\mathbf{x})$

Principled choice of α_t's: minimize the error on an "aggregation data set" pseudo-inverse

Some α_t's can come out negative

Most valuable h_t in the blend?