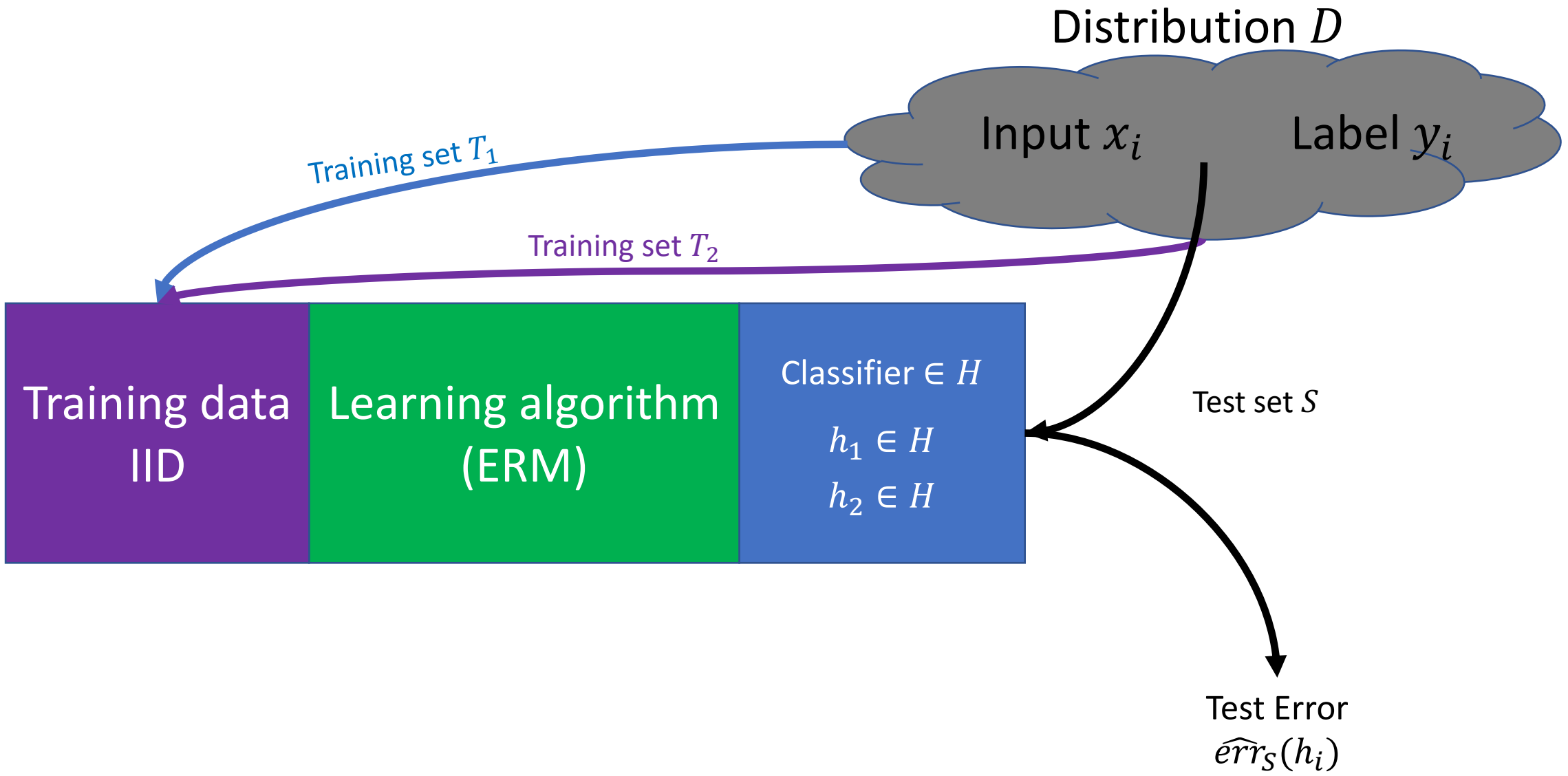
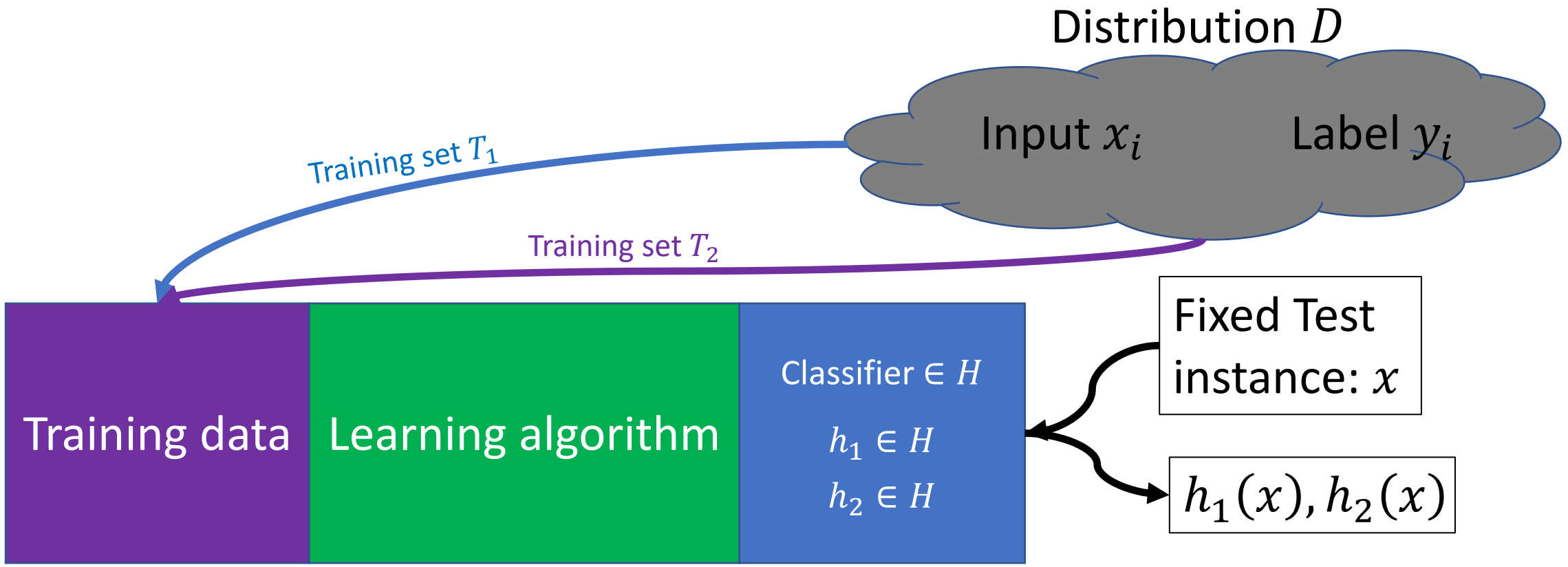
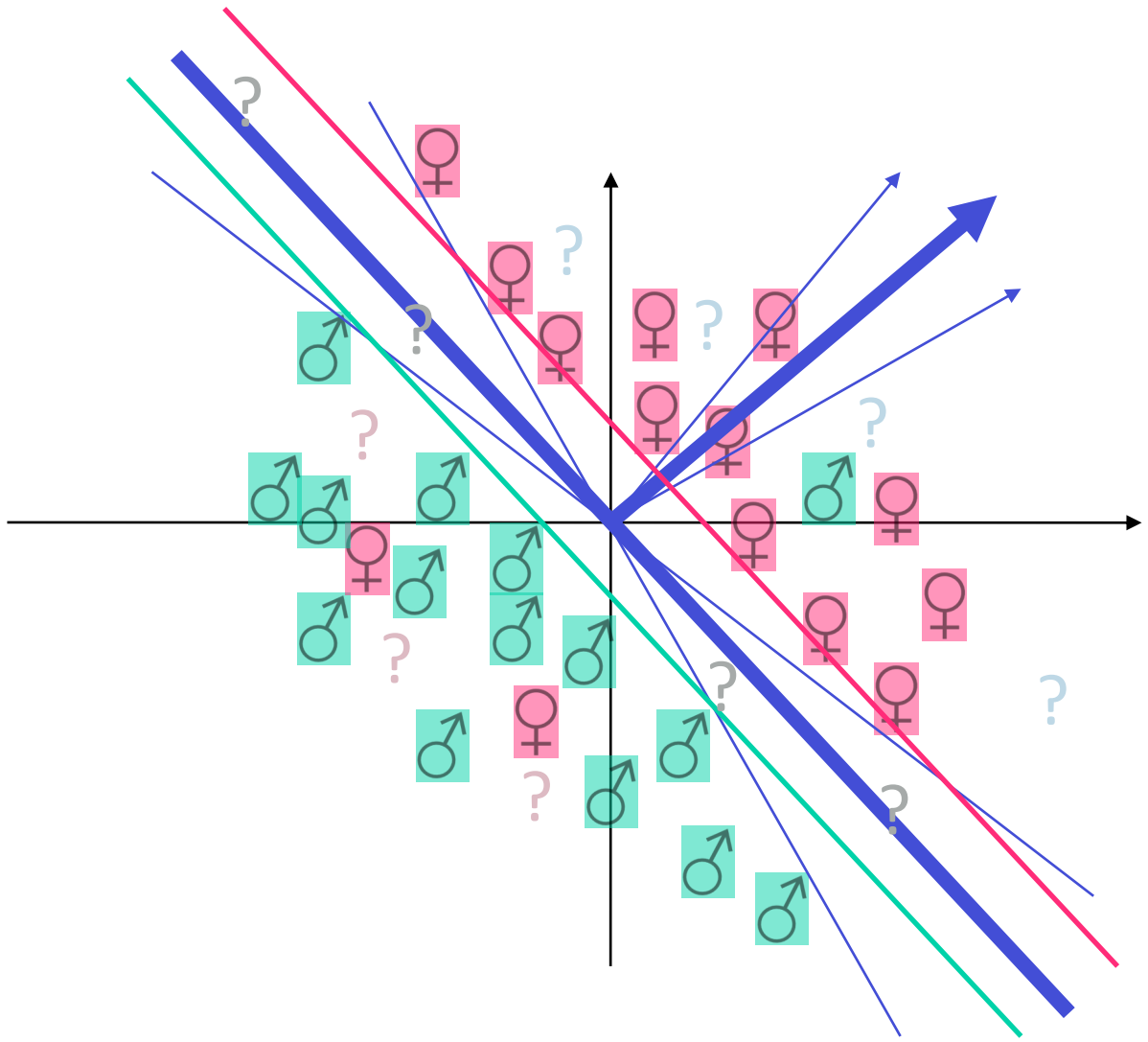

Distribution D
Input x_i
Label y_i
Training set T_1
Training set T_2
Training data
IID
Learning algorithm
(ERM)
Classifier $\in H$
$h_1 \in H$
$h_2 \in H$
Test set S
Test Error
$\widehat{err}_S(h_i)$


Distribution D
Input x_i
Label y_i
Training set T_1
Training set T_2
Training data
Learning algorithm
Classifier ∈ H
$h_1 \in H$
$h_2 \in H$
Fixed Test instance: x
$h_1(x), h_2(x)$

TLDR2: Margins and confidence



- Find linear classifier with large margin
- Large margin implies there is a **cone** of linear classifiers with close-to-minimal error.
- Large margin test examples are classified identically by classifiers in cone.
- **The prediction on large margin test examples is insensitive to small changes in training set.**

Stability wrt other perturbations

- Perturbing the data
  - margins: replacing the training set.
  - Bootstrap: approx. replacing dataset when data is limited.
  - Bagging / random forests.
  - Adding noise to the training data.
  - Getting data from different sources.
- Perturbing the algorithm:
  - Varying the model: NN architecture, depth of decition trees, number of features.
  - Combining completely different algorithms: NN, Boosting, KNN, decision trees
- Perturbing the input:
  - Adding noise (random, adversarial)
  - Transforming (for images: rotation, translation, scaling,…)

Some properties of δ-stability

- If x is δ-stable, and $\delta < \frac{1}{2}$ then by taking the majority vote over $\frac{1}{\left(\frac{1}{2}-\delta\right)^2}\log\frac{1}{\epsilon}$ hypotheses gives a rule that is ϵ-stable.
- Define the expected stability to be: $\Lambda = P_{x,h_1,h_2}[h_1(x) \neq h_2(x)]$ then $\Lambda \geq |err(h_1) - err(h_2)|$
- If $err(h_1), err(h_2) \leq \epsilon$ then …

Detection cascade