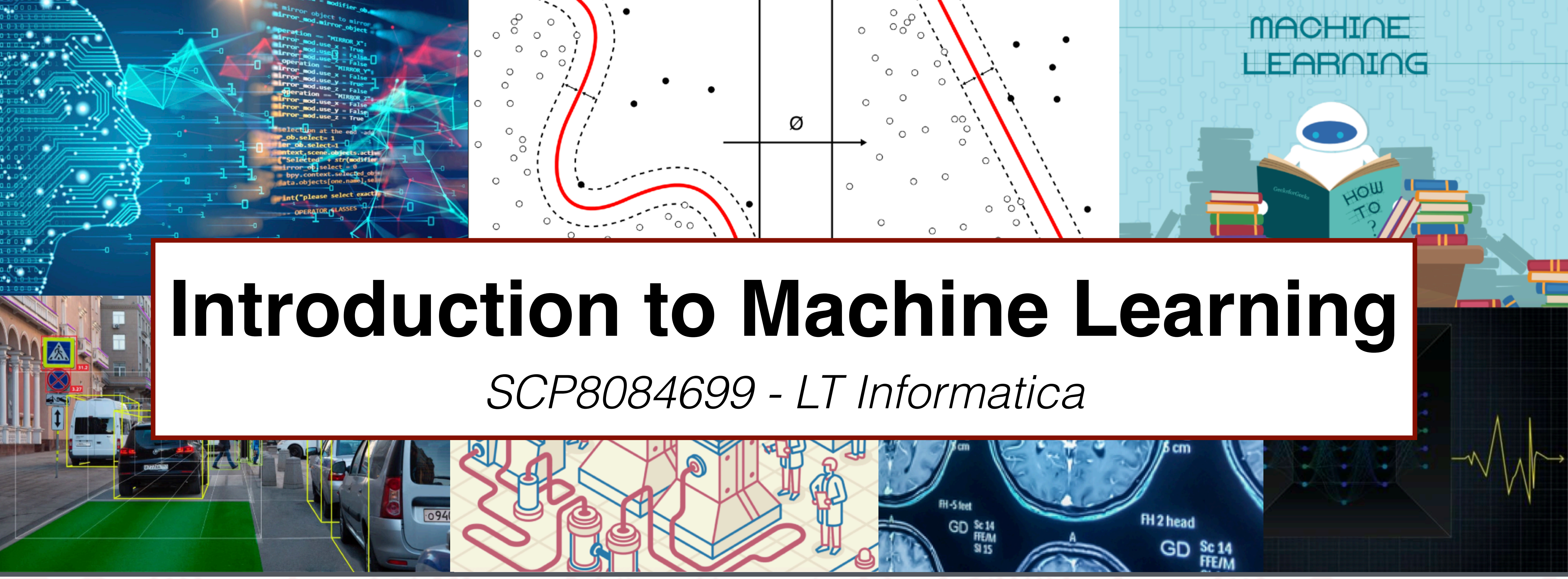


Introduction to Machine Learning

SCP8084699 - LT Informatica

Model Selection and Evaluation

Prof. Lamberto Ballan

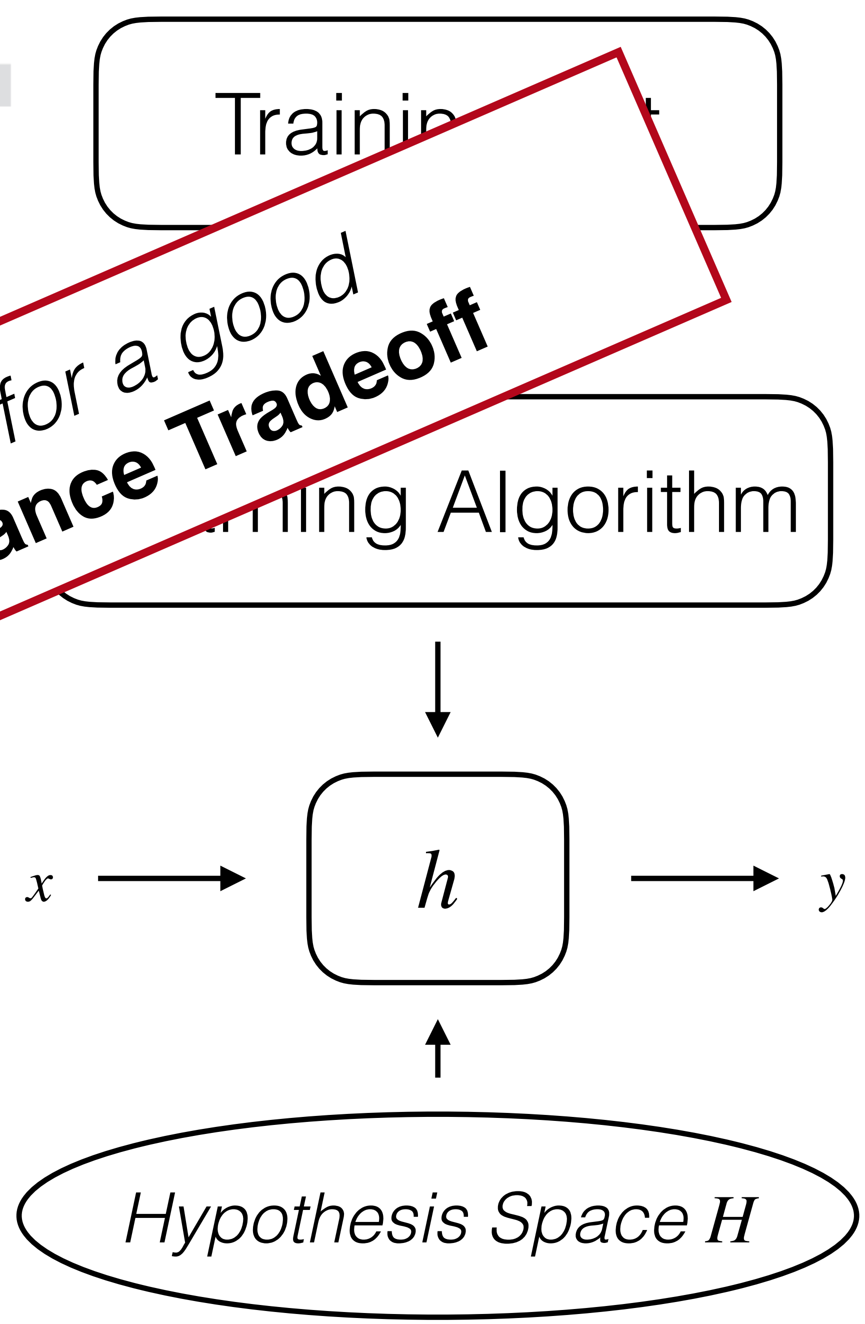
DIPARTIMENTO MATEMATICA
Department of Mathematics "Tullio Levi-Civita"

Recap: Supervised Learning

- Classification (discrete) *vs* Regression (real-valued output)

$\{(x^{(i)}, y^{(i)})\}$

Size in feet² (x)	Price ($) in 1K's (y)
2104	460
1416	232
1534	315
852	178
…	…



Looking for a good
Bias-Variance Tradeoff

h approximates the unknown target f

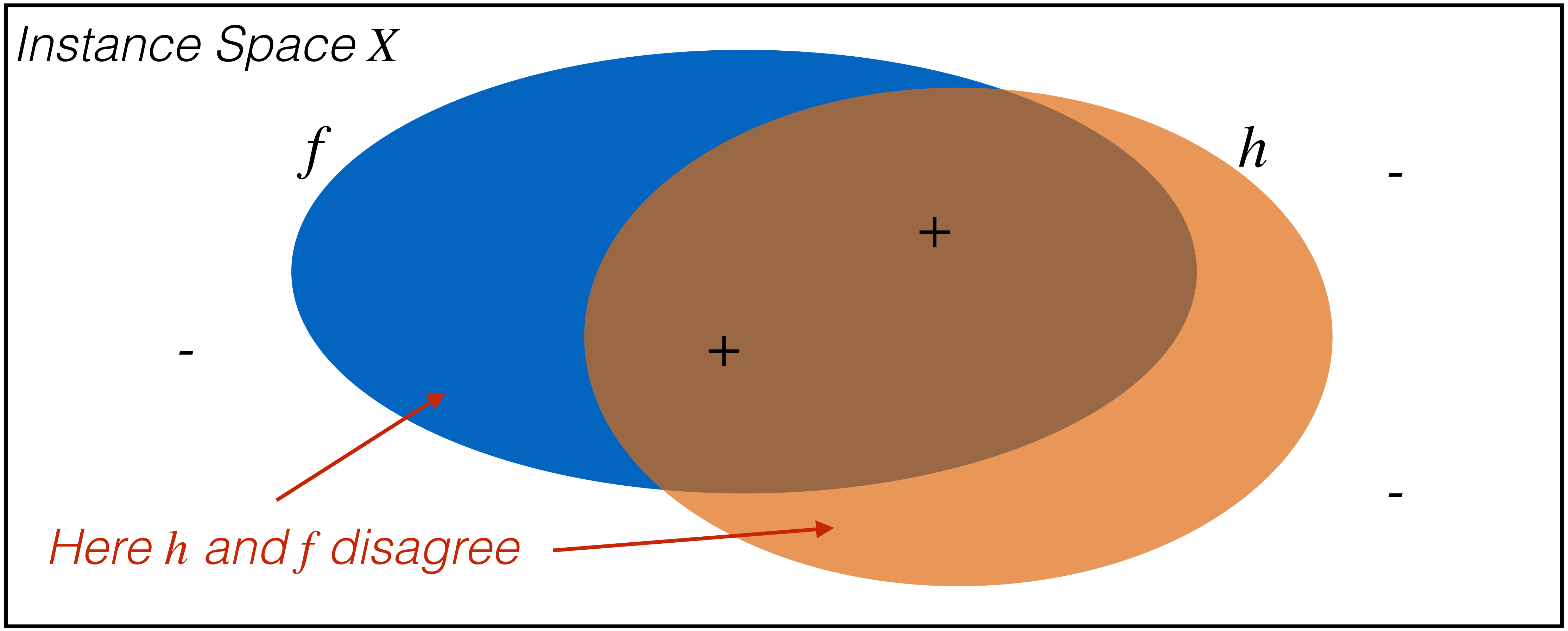
$$h \sim f: X \to Y$$

The “Real” Scenario

- Let me know introduce a more formal definition for a machine learning problem
- **Empirical Risk Minimization**: it defines a family of learning algorithms and is used to give theoretical bounds on their performance
    - We don’t know how well a learning algorithm will work in practice (“true risk”) because we don’t know the true distribution of data
    - But we can measure its performance on our training dataset (“empirical risk”)

True Error

$h \sim f: X \rightarrow Y$

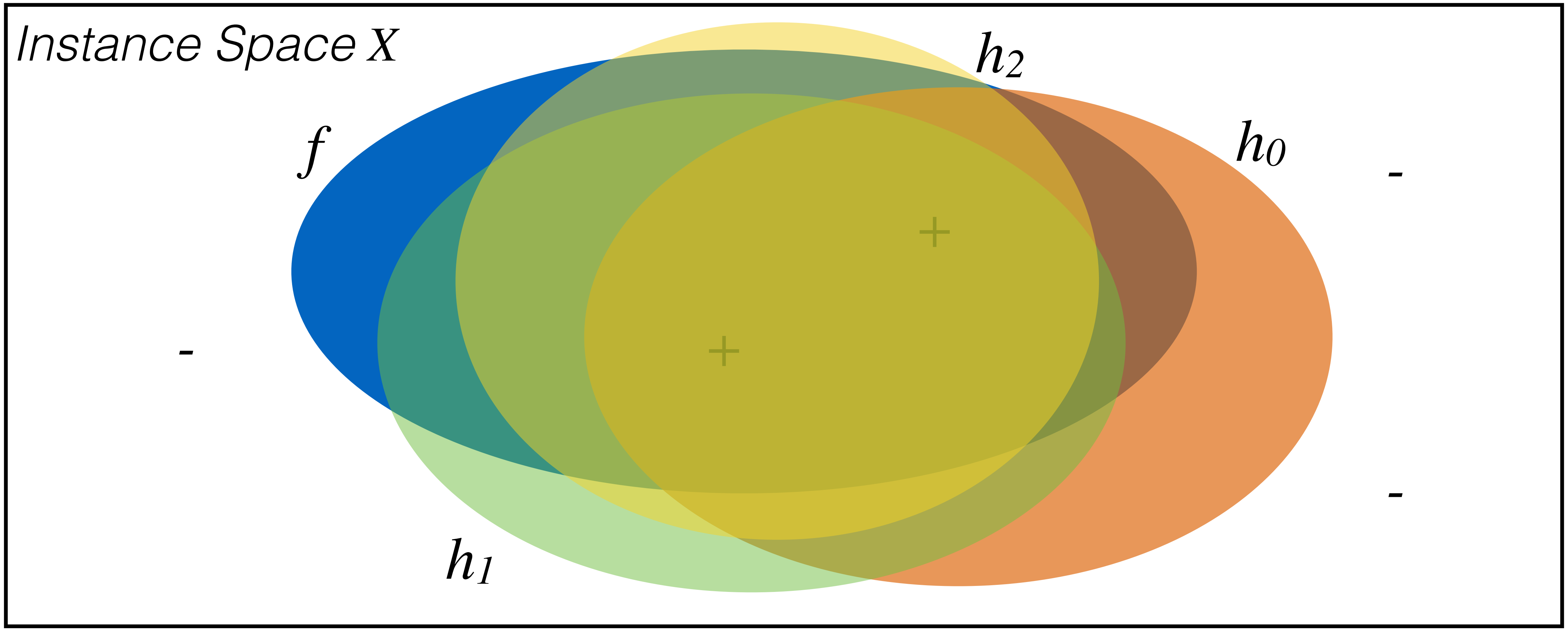


True Error of hypothesis h with respect to target f and distribution D (to observe an input instance $x \in X$) is the probability that h will misclassify an instance drawn at random according to D:

$$error_D(h) \equiv \underset{x \in D}{\text{Prob}}\{\, f(x) \neq h(x) \,\}$$

Empirical Error

$h \sim f: X \rightarrow Y$



Empirical Error of hypothesis h with respect to a training set T is the number of examples that h will misclassifies:

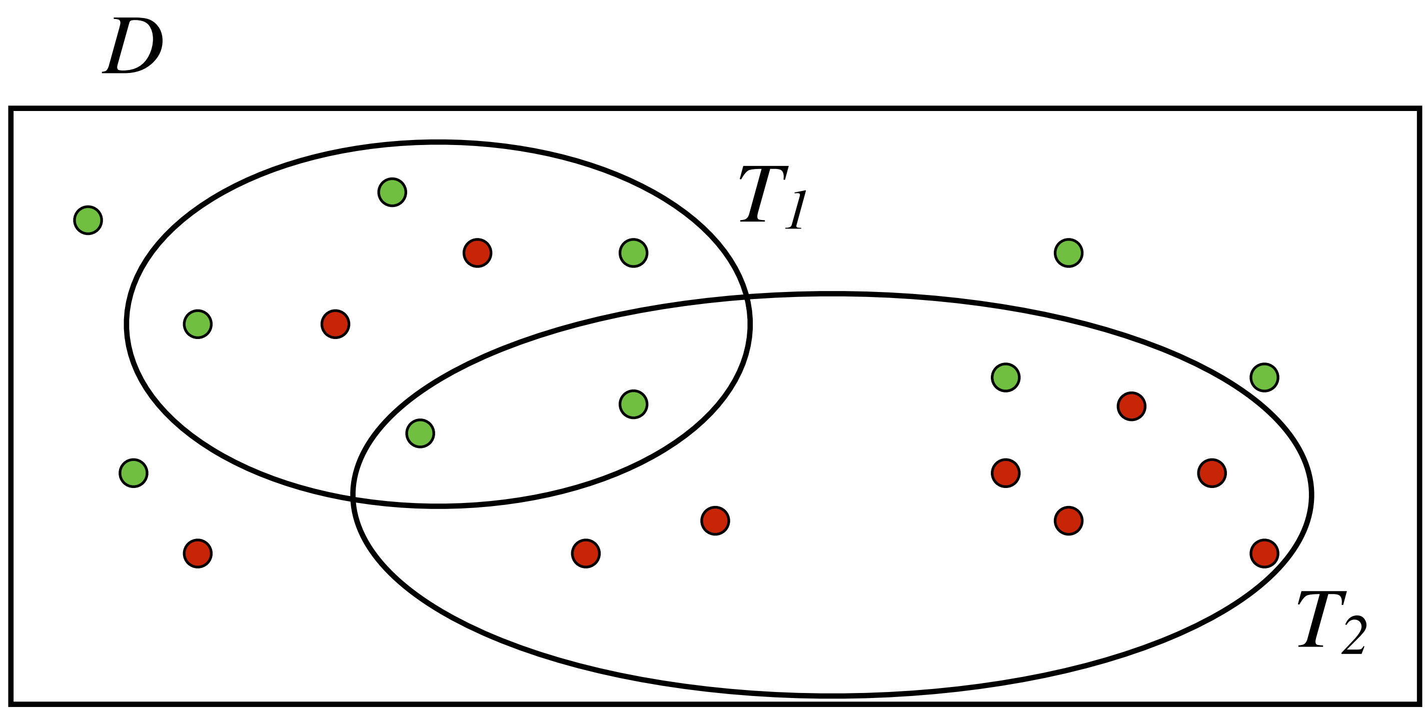
$$error_T(h) \equiv \#\{\ (x, f(x)) \in T \mid f(x) \neq h(x)\ \}$$

Training (sample) Error

True vs Empirical Error

- A simple example:

  - Correctly classified: $f(\mathrm{x}) = h(\mathrm{x})$
  - Misclassified: $f(\mathrm{x}) \neq h(\mathrm{x})$



$$error_D(h) = 0.5$$

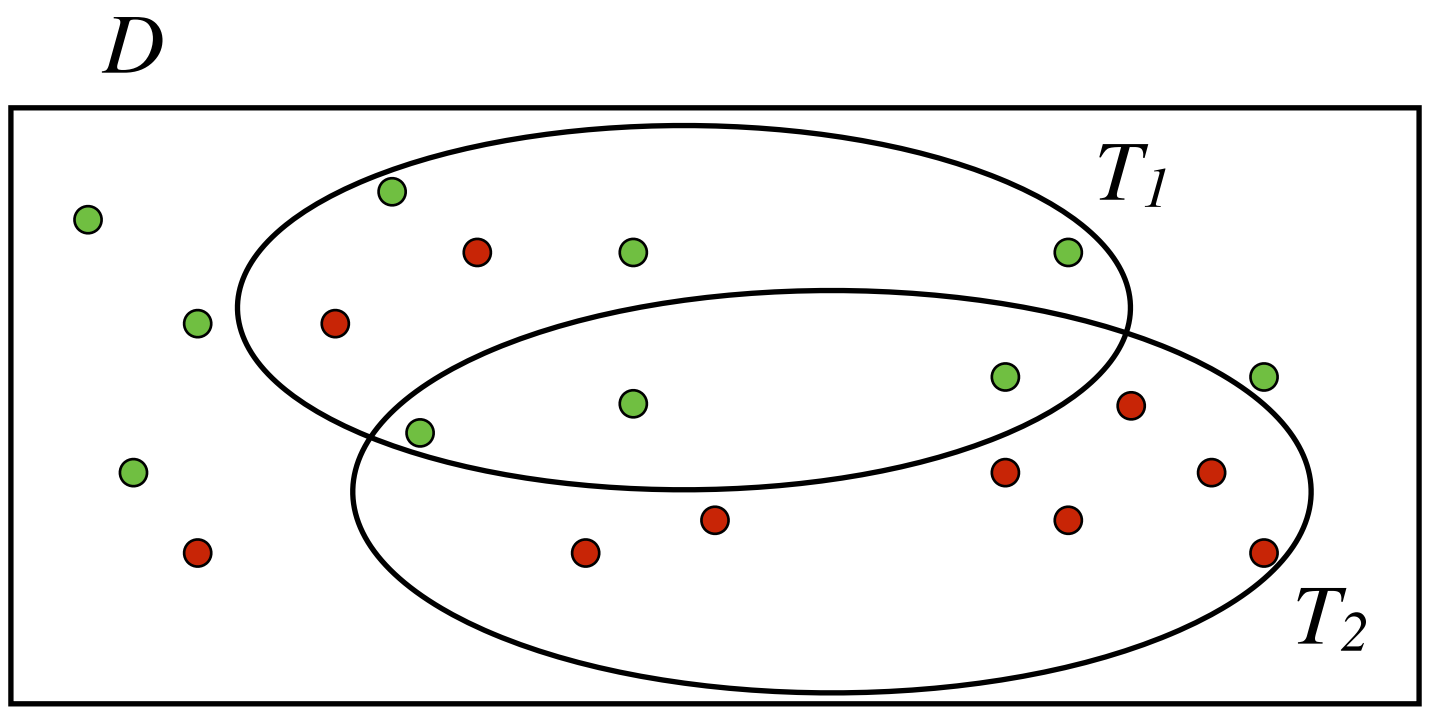
$$error_{T_1}(h) = 0.29$$

$$error_{T_2}(h) = 0.7$$

- Generally, we never know the true error; we only get to see the empirical error.
- How well does the empirical error estimate the true error?

Bias-Variance Tradeoff

- Let's now look at Bias and Variance:



$$error_D(h) = 0.5$$

$$error_{T_1}(h) = 0.25$$

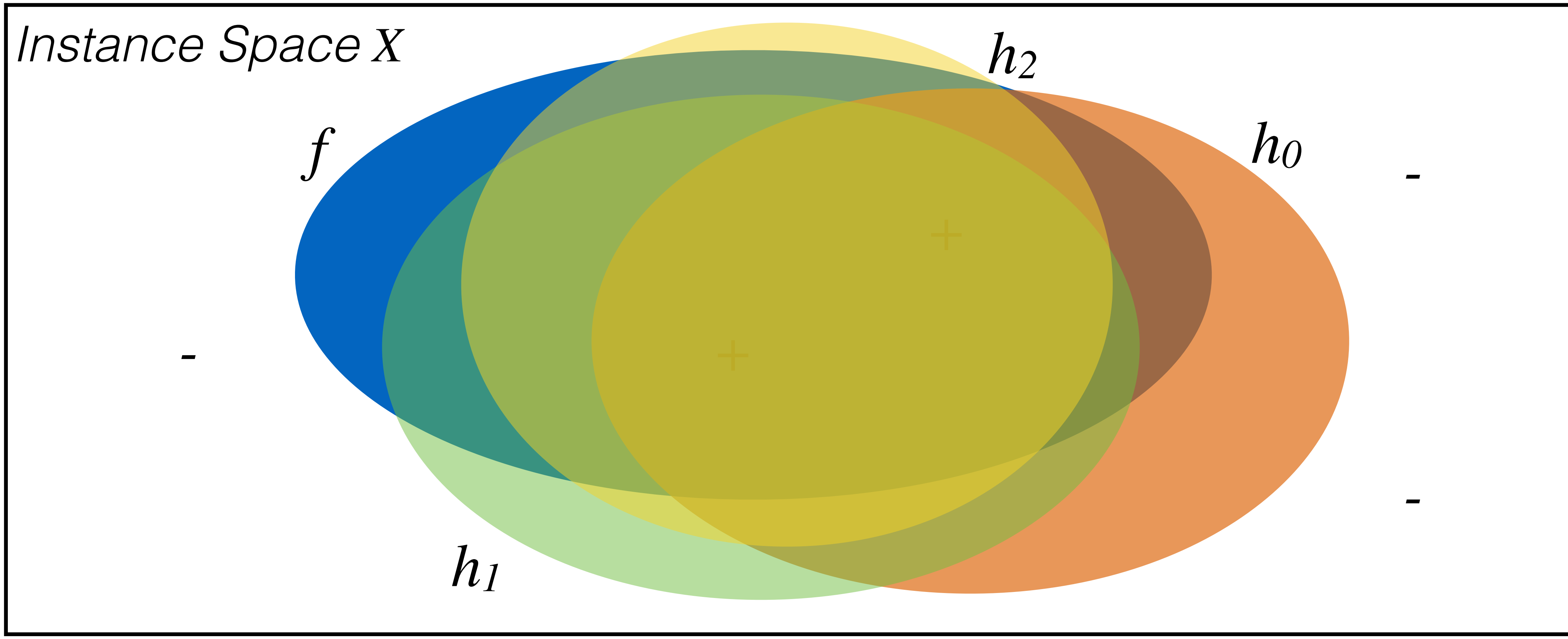
$$error_{T_2}(h) = 0.7$$

- **Bias:** here $error_{T_1}(h)$ is optimistically biased

- **Variance:** even without bias, $error_{T_j}(h)$ may still vary from $error_D(h)$

Bias-Variance Tradeoff

$h \sim f: X \rightarrow Y$



Overfitting: given $h \in H$, h overfits T if $\exists h' \in H$ such that $error_T(h) <$ $error_T(h')$ but $error_D(h) > error_D(h')$

Model Selection and Evaluation

- **Hold-out**: we keep a subset of v samples from the training set (the validation set) to evaluate our model
  - A classifier/regressor is trained on m-v samples
  - Parameters are optimized on the training-validation sets: then you should evaluate performances on the test set
  - Size (cardinality) of training+validation sets should be greater than test set, e.g. 70%,15%,15%

Model Selection and Evaluation

- ***k*-fold cross validation**:
    - k classifiers/regressors $h_1, \ldots, h_k$ are trained by partitioning the training set T into k disjoint sets (folds) $V_1, \ldots, V_k$, and then iteratively applying the hold-out approach
    - Overall performance is measured averaging the results obtained on the different folds
- **Leave-one-out**:
    - It is a special case of *k*-fold cross validation in which k is equal to the cardinality of the training set

Model Selection and Evaluation

- ***k*-fold cross validation**: an example (*5*-fold)

Metrics

- How to evaluate performances?
  - We will look at different metrics for classification tasks
  - Let's start with a simple example: dog vs no-dog

TP: True Positive **FP:** False Positive **TN:** True Negative **FN:** False Negative

Metrics

- How to evaluate performances?

$$\textbf{Accuracy} = \frac{TP + TN}{P + N} = \frac{\textit{all correct}}{\textit{all instances}} = \frac{4 + 4}{11} = 0.73$$

TP: True Positive **FP:** False Positive **TN:** True Negative **FN:** False Negative

Metrics

- How to evaluate performances?
  - **Confusion Matrix**:

		Actual Class (groundtruth)	
		dog	no-dog
Predicted	dog	**TP:** True Positive	**FP:** False Positive *Type I error*
	no-dog	**FN:** False Negative *Type II error*	**TN:** True Negative
		P	**N**

Contact

- **Office:** Torre Archimede, room 6CD3
- **Office hours** (ricevimento)**:** Friday 9:00-11:00

✉ lamberto.ballan@unipd.it

⌂ http://www.lambertoballan.net

⌂ http://vimp.math.unipd.it

@ twitter.com/lambertoballan