

Statistical Learning Theory

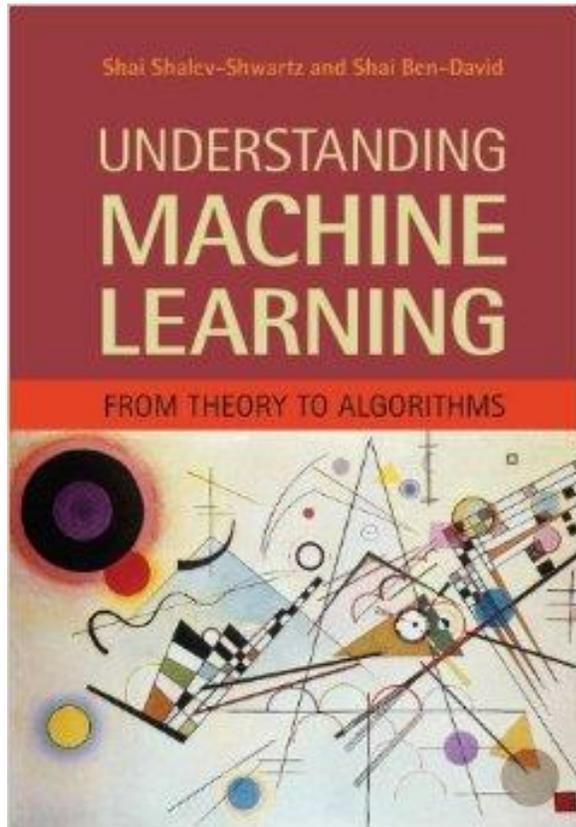
Introduction

Course Details

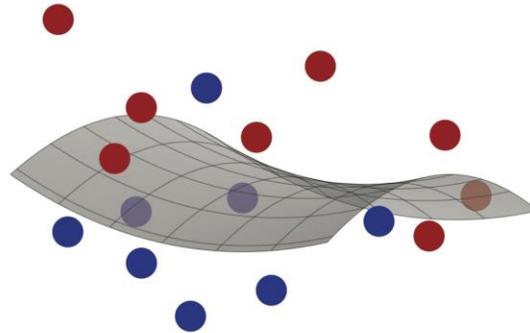
- Pabitra Mitra – (pabitra@cse.iitkgp.ac.in) (Instructor)
- Dhruv Sarkar (dhruv.sarkar223@gmail.com) (Teaching Assistant)
- Smarak Kanjilal (smarakkanjilal@gmail.com) (Teaching Assistant)

- Course Webpage:
<https://cse.iitkgp.ac.in/~pabitra/course/aml/aml.html>

Textbooks



Foundations of Machine Learning second edition



Mehryar Mohri,
Afshin Rostamizadeh,
and Ameet Talwalkar

Syllabus

- Learnability, PAC Learning, Uniform Learnability
- Vapnik-Chervonenkis dimension, Growth Functions
- Structural Risk Minimization
- Weak Learnability and Boosting
- Online Learning Models

- TBD....

Evaluation

- Midsem – 30%
- Endsem – 50%
- Scribe-notes (Teacher Assessment) – 20% (4 lectures per student)
 - LaTeX file to be submitted in moodle within one week of class
- Term Project (optional and zero credit)

An example learning problem

- Predict if a GUAVA is TASTY.
- Based on its COLOR and SOFTNESS
- You can actually taste (after buying) n number of Guavas and find out if they are tasty.
- Learn from this experience.
- Predict for a new Guava before buying....

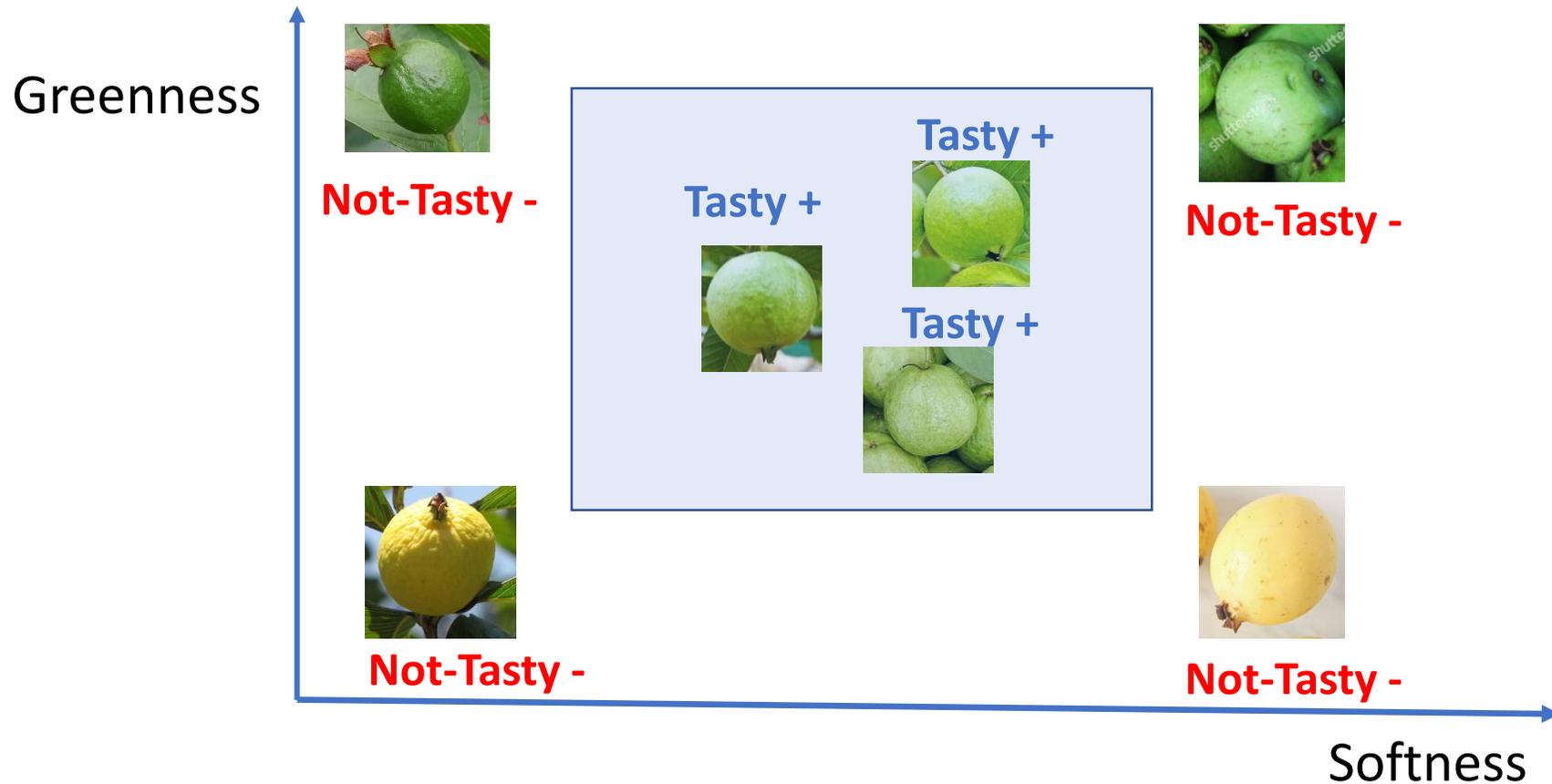
Representing Guavas



Tasty Guavas

- Not too Green - Not too Yellow
- Not too Soft - Not too Hard

Hypothesis: Tasty Guavas – Axis Parallel Rectangle



Indicator Function F – Axis Parallel Rectangle (Fixed but Unknown)

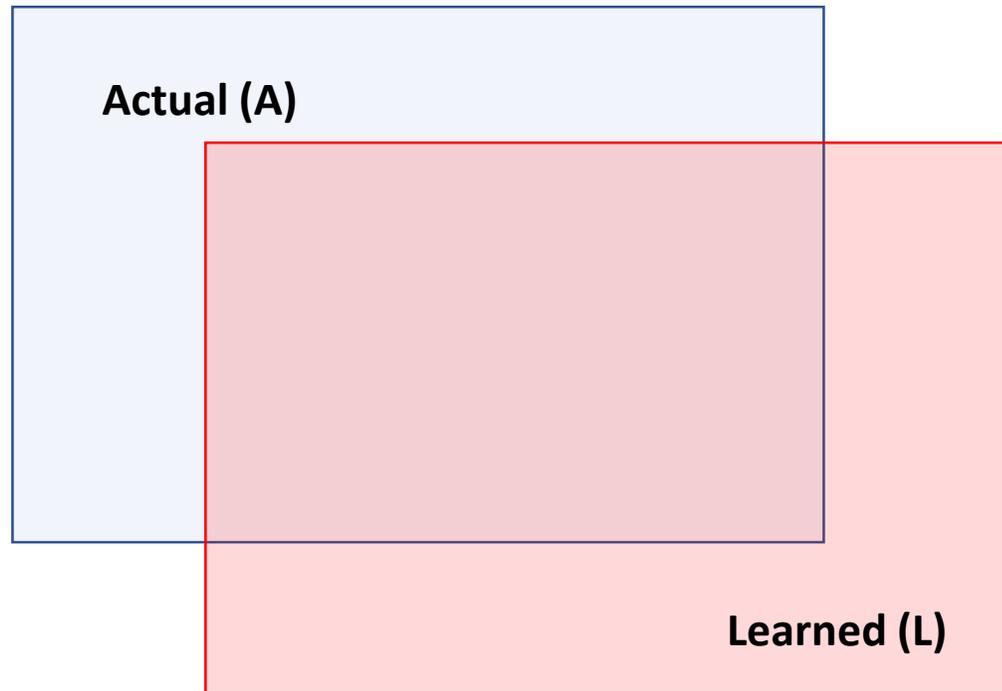
Learning Problem

- Guess the unknown axis-parallel rectangle after seeing n examples
 - Examples are independent and identically distributed (i.i.d)
 - Drawn from the same distribution as data
 - Fixed but unknown data distribution

Learning Algorithm

- Return the smallest rectangle containing all the positive examples out of the n samples

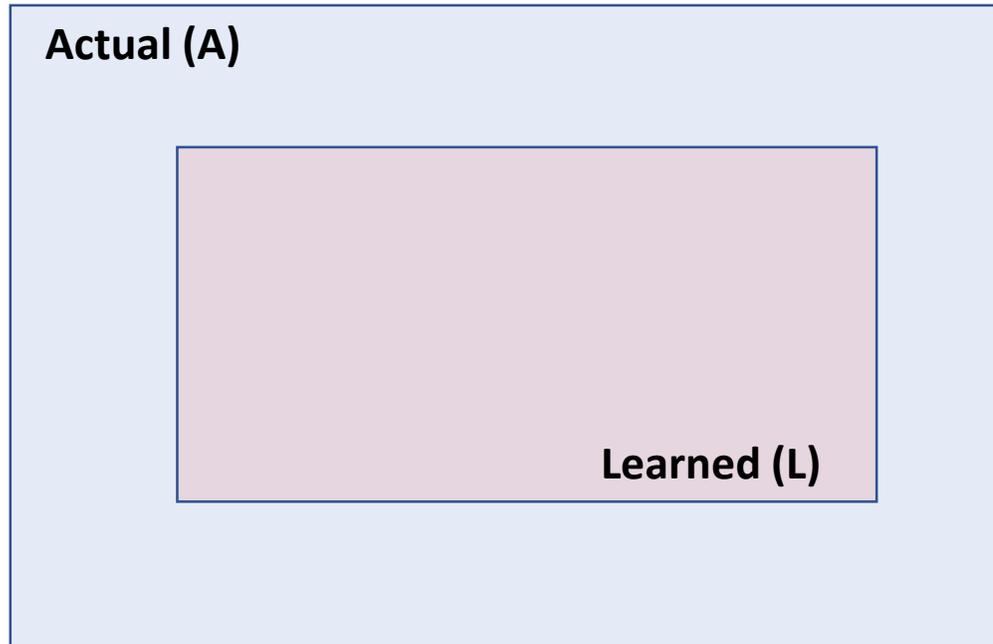
Learning Error



$$\text{Error} = (A - L) \cup (L - A)$$

(Symmetric Difference)

Error (ϵ) for our Algorithm



Area of the region between two rectangle

– Probability of misclassification on a random new Guava

Question (trivial)

- What is the value of n such that ε is 0 ?

Answer

- What is the value of n such that ε is 0 ?
 - *Infinity*

Question

- What is the error ε after looking at n examples?

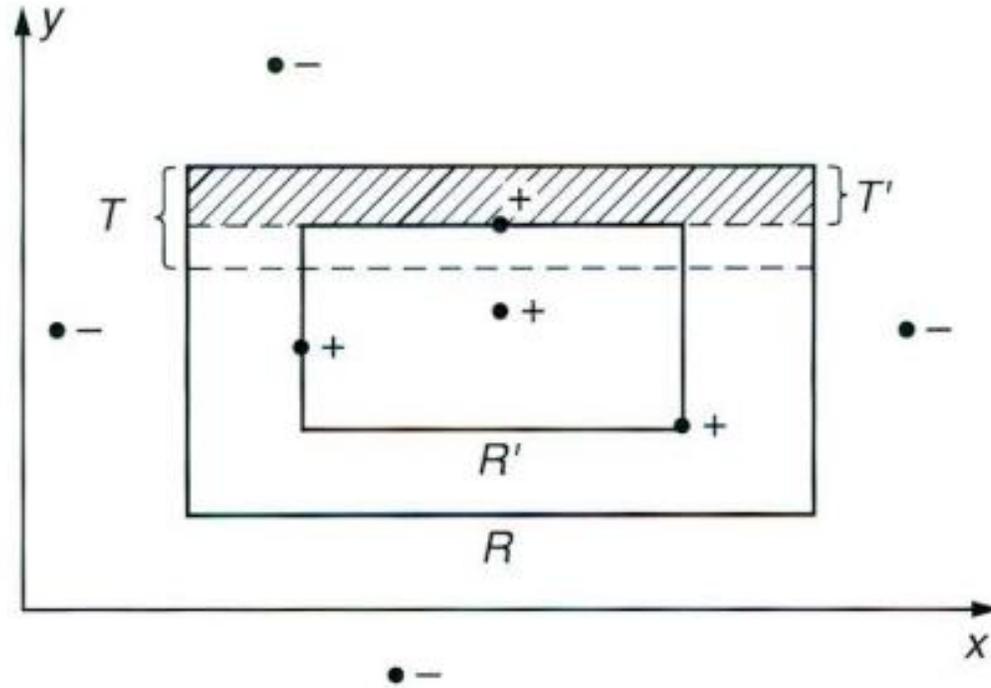
Answer

- What is the error ε after looking at n examples?
 - Depends on the set of examples used for training

Modified Question

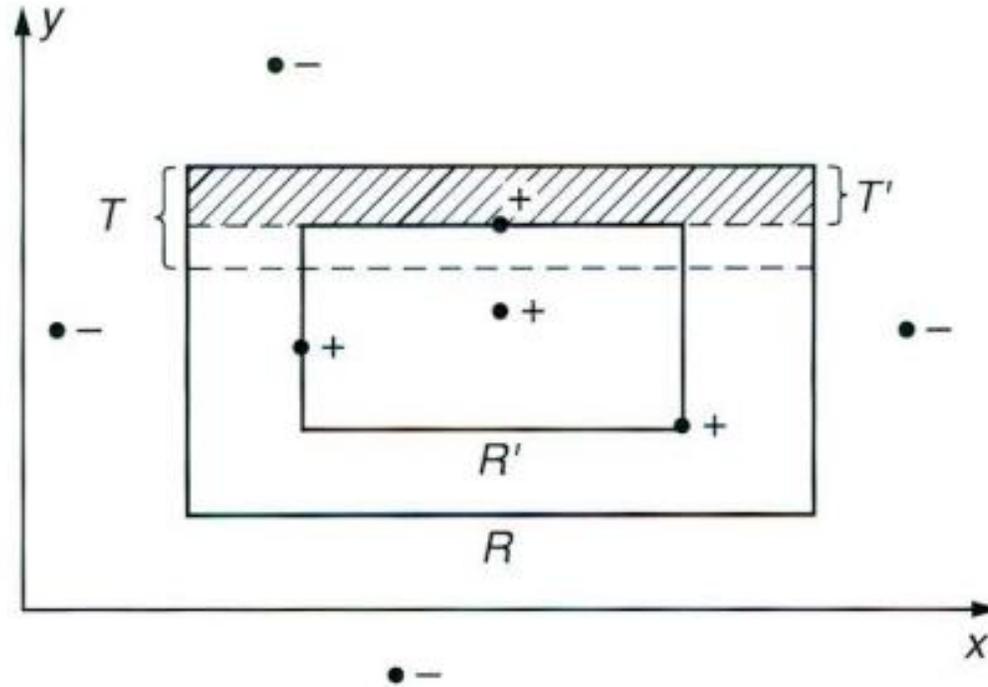
- What is the probability δ that the error is $\leq \epsilon$, for a given ϵ , after looking at n examples?
 - iid samples
 - Fixed but unknown distribution of data D
 - Distribution D can be anything (worst case analysis)

Analysis



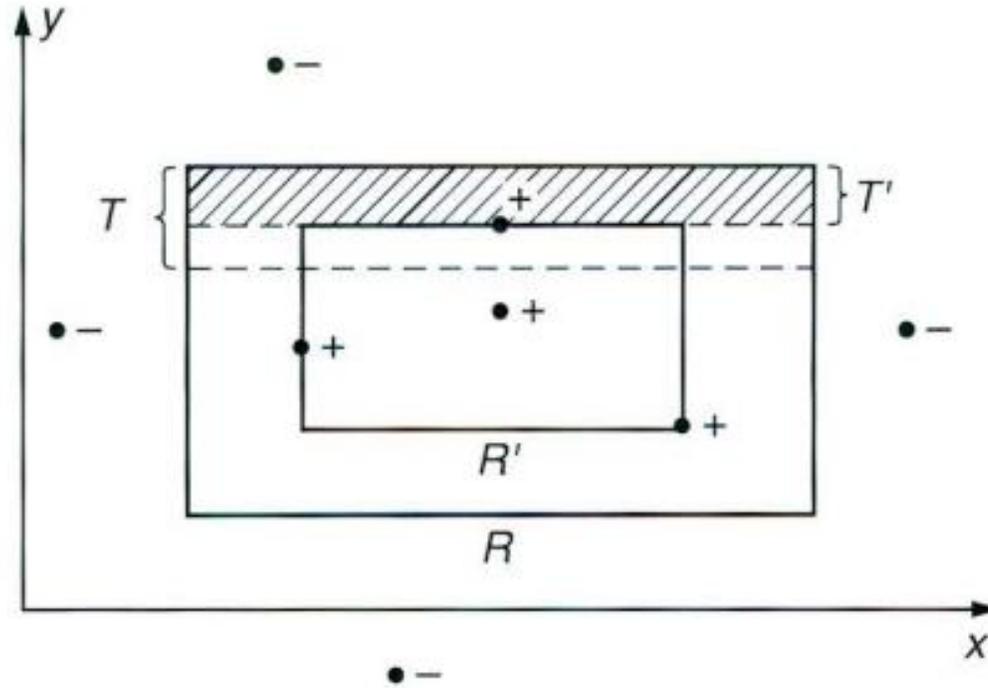
Strip T has a weight $\varepsilon/4$ under D

Analysis



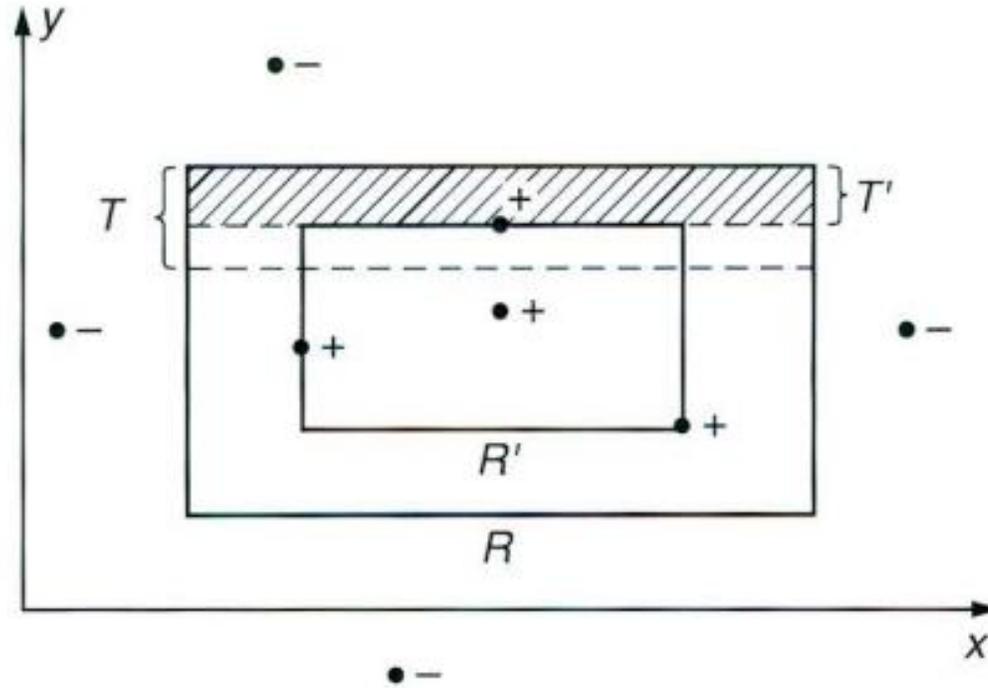
T' is smaller than T ($\epsilon/4$) if one of the n examples lie in T .

Analysis



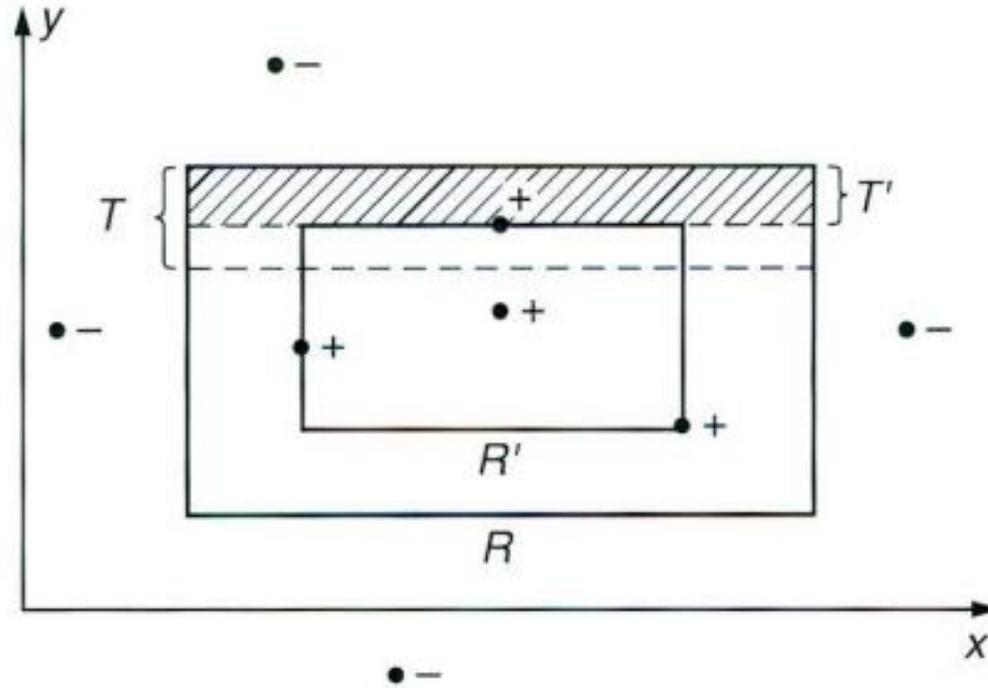
What is the probability that an example lies in T ?

Analysis



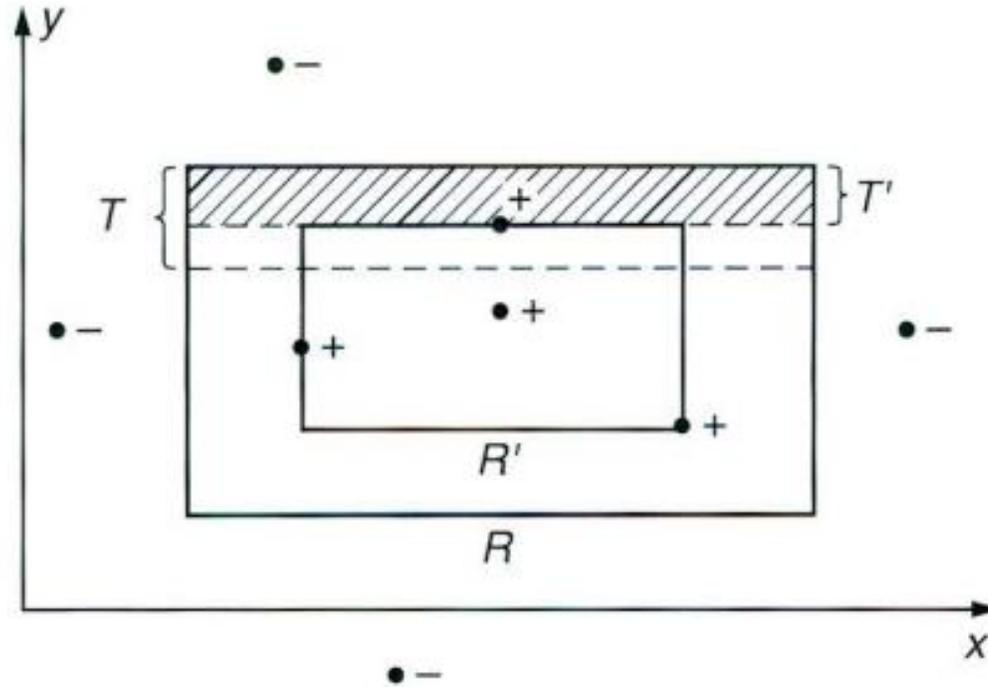
The probability that an example lies in $T = \epsilon/4$

Analysis



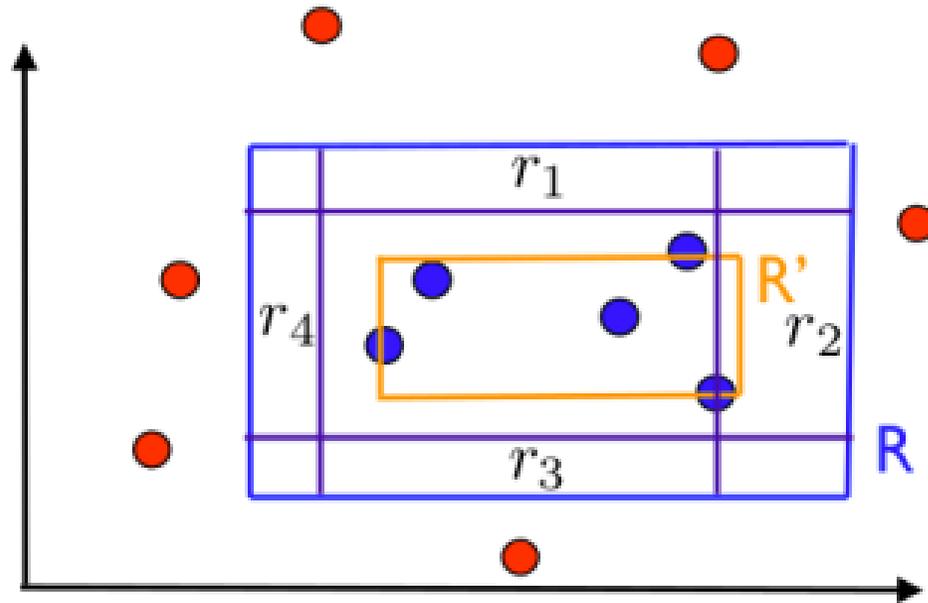
The probability that an example does not lie in $T = 1 - \epsilon/4$

Analysis



The probability that none of the n i.i.d. example lie in $T = (1 - \epsilon/4)^n$

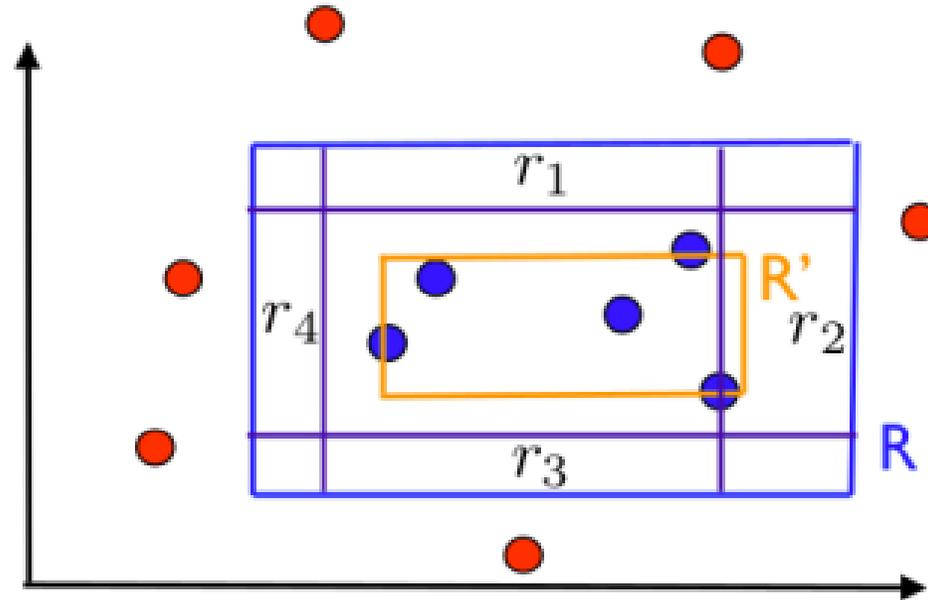
Analysis



The probability that none of the n i.i.d. example lie in all four strips = $4(1 - \epsilon/4)^n$

$$Pr[A \cup B] \leq Pr[A] + Pr[B] \quad (\text{Union Bound})$$

Analysis



If we want $4(1 - \epsilon/4)^n$ that $\leq \delta$, then $n \geq (4/\epsilon) \ln(4/\delta)$ (sample complexity bound)

$$(1 - x) \leq e^{-x}$$