Learning theory

inductive bias

first of all: learning without inductive bias is impossible!!

linear regression

the meaning of the coefficient b_i in

$$Y = a + b_1X_1 + b_2X_2 + \cdots + b_mX_m$$

b_i tells us how much increases if X_i increases by 1 while all X_j, j≠i, remain the same -> the effect of X_i on Y

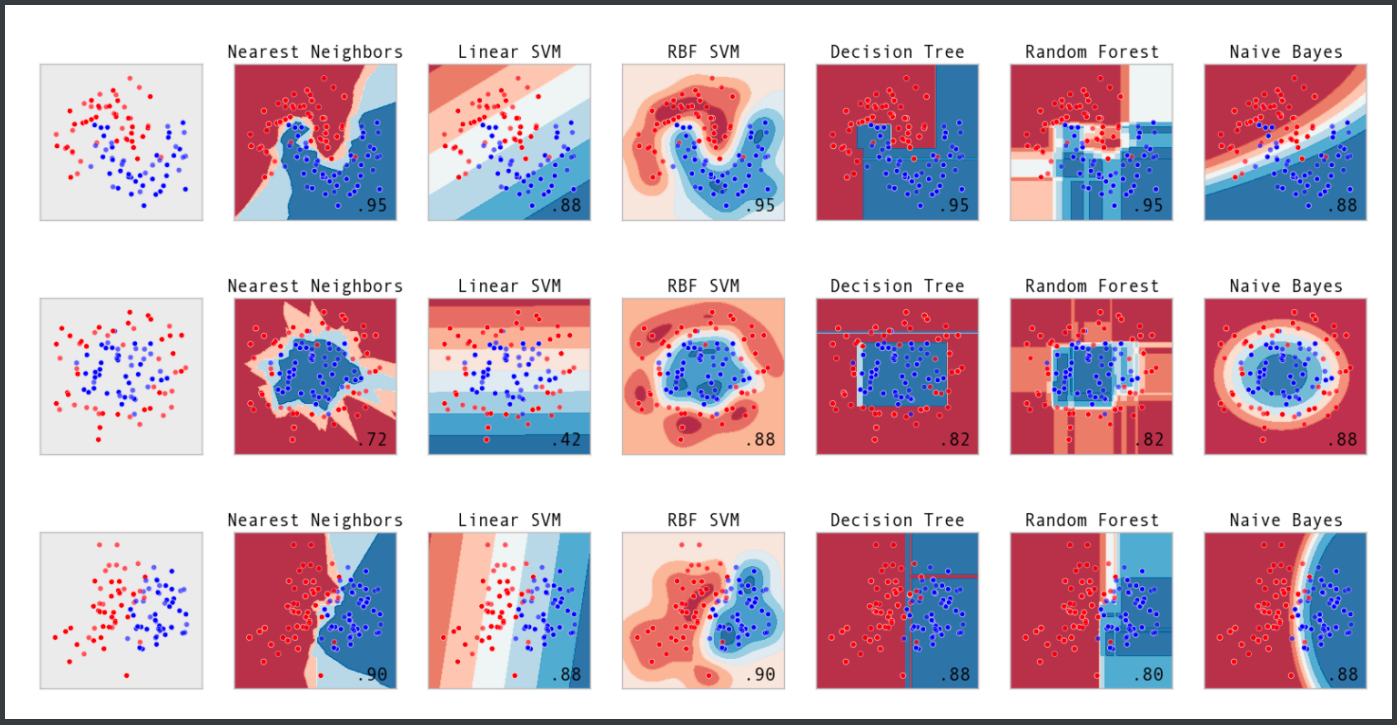
- Watch out - Not **causal**! we just **EXPECT** Y to increase
- no interactions among X_i's (constant effect & cumulative)

decision tree

not anymore "X_j has the same effect everywhere",

X_j might have a positive effect on Y in one branch, but a negative in another

Bias: different methods, different implicit assumptions



no generalization towards unseen cases is possible without making some assumptions

- how can we measure the strength of the bias

  generalizability
- how does bias impact learnability

  negative of strong bias:

  circumstances where the model could be used is limited

  positive of strong bias:

  learning would be easier

learning formal langugaes & automata

decision problems -> formal languages:

- **Question**: under what conditions is the following possible: given a set of strings that do (not) belong to a language L, construct a machine that "decides L" (=

identifiable in the limit:

其中T是目标语言，L是猜测，只要有inconsistent的样本就持续更新L；

- Definition: A class of languages $\mathscr{C}$ is **identifiable in the limit** if a learner exists with the property that, for any $T \in \mathscr{C}$, L will necessarily become equal to T after a finite number of examples.

（即，学习器能够通过有限数量的样本稳定在正确的语言上）

- positives only, even relatively simple classes of languages (e.g. regular languages) are not identifiable in the limit
- positives & negatives, most classes are identifiable in the limit

RPNI 算法：given p/n examples, to construct a FSA

step 1: construct the PTA (prefix tree acceptor)

- 为每个 positive 字符串提供唯一的路径，不接受其他字符串；
- 此时满足：接受所有positive，拒绝所有negative；

step 2: iteratively merge

- 目的：generalize FSA，使其接受更多串
- 合并规则：

  如果 $\delta(q_i, \sigma) = q$ 且 $\delta(q_j, \sigma) = q'$ 那么必须将 q 和 q′ 合并，同时 $\delta(q_{i,j}, \sigma) = q$；

  如果合并前 q 和 q′ 有任何一个是合并状态，则合并之后也必须是合并状态；（这一步让FSA保持接受原来接受的串，同时接受更多串）
- 合并顺序：

  红色节点不能合并红色，蓝色节点是红色的后继，蓝色如果不能再合并则变红

---

measuring the expressiveness of a hypothesis space

PAC learning