

SIGEVO Summer School 2017 Challenge

Challenge posed by Dr John McCall

Shashank Srikant

July 19, 2017

shash@mit.edu

What is the real-world relevance of problem features abstracted by machine learning?

- In an output prediction task typical of machine learning, we start with knowledge/observations, Θ of something we possess in order to predict a property, P .
- Since we are unaware of the behavior of P , we utilize the characteristics of Θ to describe P .
- We generally map what we know of Θ and P onto the real line and define a function f such that $f(\Theta) = P$. Machine learning involves learning this function f

Response (contd..)

- We define Θ to be the **features** which help describe our desired property P .
- Without knowing about P , we are able to infer from the features what P would look like. They help in predicting the structure of P .
- Hence, **features**, each of which is generally represented as a vector of numbers, are linearly or non-linearly combined to predict P
- Some examples of such features could be
 - In order to predict a user is likely to buy a product she has viewed on a web page (property P), the number of seconds stayed on the page (f_1), number of weblinks clicked (f_2), the time of the day (f_3), could possibly be features which could help predict it.
 - In order to determine the damage caused by wind in forests, a number of features like the mean tree height, density of the forest etc. could predict such damage in advance.