Predictive Analytics Lecture 1

Adam Kapelner

Stat 422/722
at The Wharton School of the University of Pennsylvania

January 17 & 18, 2017

Define: Prediction and Forecast

"statement about an uncertain event",

Define: Prediction and Forecast

"statement about an uncertain event", "informed guess or opinion"

predict (v.) 1620s (implied in predicted), "*foretell, prophesy*," a back formation from prediction or else from Latin praedicatus, past participle of praedicere "foretell, advise, give notice,"

forecast (n.) early 15c., "*forethought, prudence*," probably from forecast (v.). Meaning "conjectured estimate of a future course" is from 1670s.

I will be using predict and forecast interchangeably.

Examples

We make predictions all the time, sayomg things like:

- "Apple stock will go up tomorrow",
- "This condo will sell for $500K"

sometimes unknowingly

- "Going skiing this weekend will make me happy",

How do we make predictions?

Examples

We make predictions all the time, sayomg things like:

- "Apple stock will go up tomorrow",
- "This condo will sell for $500K"

sometimes unknowingly

- "Going skiing this weekend will make me happy",

How do we make predictions? We use a *model*.

Define: model

"a functional decription of a system"

Define: model

"a functional decription of a system"

An example model is:

> *Early to bed and early to rise make a man healthy, wealthy and wise.*

All models have **input(s)** that are then somehow related to its **output(s)**. What are the input(s) and output(s) above?

Inputs:

Define: model

"a functional decription of a system"

An example model is:

> *Early to bed and early to rise make a man healthy, wealthy and wise.*

All models have **input(s)** that are then somehow related to its **output(s)**. What are the input(s) and output(s) above?

Inputs: bedtime schedule, waking schedule
Outputs:

Define: model

"a functional decription of a system"

An example model is:

> *Early to bed and early to rise make a man healthy, wealthy and wise.*

All models have **input(s)** that are then somehow related to its **output(s)**. What are the input(s) and output(s) above?

Inputs: bedtime schedule, waking schedule
Outputs: health, wealth and wisdom

Observations and their Features

Here, the inputs and outputs are

- *features* or
- *attributes* or
- *characteristics* of

a person.

Generally, inputs and outputs are features of the

- *unit of analysis* or
- *the observation* or
- *the subject*.

Thus the model relates some *feature(s) of the observation* to other *feature(s) of the observation*. Here, we are relating specific people's bedtime schedule and waking schedule to their health, wealth and wisdom.

Features as Measurements

Models phrased in language such as:

> *Early to bed and early to rise make a man healthy, wealthy and wise.*

are usually ambiguous, imprecise, vague and ill-defined. Why?

Features as Measurements

Models phrased in language such as:

> *Early to bed and early to rise make a man healthy, wealthy and wise.*

are usually ambiguous, imprecise, vague and ill-defined. Why? What does "early to bed" mean? What does "early to rise" mean? What does "healthy" mean? What does "wealthy" mean? What does "wise" mean? Without resolving these ambiguities,

Features as Measurements

Models phrased in language such as:

> *Early to bed and early to rise make a man healthy, wealthy and wise.*

are usually ambiguous, imprecise, vague and ill-defined. Why? What does "early to bed" mean? What does "early to rise" mean? What does "healthy" mean? What does "wealthy" mean? What does "wise" mean? Without resolving these ambiguities, the model is unusable and of course, untestable.

Features as Measurements

Models phrased in language such as:

> *Early to bed and early to rise make a man healthy, wealthy and wise.*

are usually ambiguous, imprecise, vague and ill-defined. Why? What does "early to bed" mean? What does "early to rise" mean? What does "healthy" mean? What does "wealthy" mean? What does "wise" mean? Without resolving these ambiguities, the model is unusable and of course, untestable.

In order to make this precise and defined, there is a necessity to use numbers.

Features as Measurements

Models phrased in language such as:

> *Early to bed and early to rise make a man healthy, wealthy and wise.*

are usually ambiguous, imprecise, vague and ill-defined. Why? What does "early to bed" mean? What does "early to rise" mean? What does "healthy" mean? What does "wealthy" mean? What does "wise" mean? Without resolving these ambiguities, the model is unusable and of course, untestable.

In order to make this precise and defined, there is a necessity to use numbers. Thus, features of the observation must be *measured* (to be defined later).

Features as Measurements

Models phrased in language such as:

> *Early to bed and early to rise make a man healthy, wealthy and wise.*

are usually ambiguous, imprecise, vague and ill-defined. Why? What does "early to bed" mean? What does "early to rise" mean? What does "healthy" mean? What does "wealthy" mean? What does "wise" mean? Without resolving these ambiguities, the model is unusable and of course, untestable.

In order to make this precise and defined, there is a necessity to use numbers. Thus, features of the observation must be *measured* (to be defined later). Thus the model relates some *measured feature(s) of the observation* to other *measured feature(s) of the observation*.

The Model as a Functional Relationship

The relationship is a function taking in inputs (within the parentheses) and "returning" the outputs (the equal sign). For any observation that is:

$$\begin{pmatrix}\text{the measured}\\ \text{outputs of an}\\ \text{observation}\end{pmatrix} = \text{model}\left(\begin{pmatrix}\text{the measured}\\ \text{inputs of an}\\ \text{observation}\end{pmatrix}\right)$$

It is traditional to put the outputs on the left hand side. This is assumed that the outputs were measured. This type of observation is called

- old
- historical
- known

In our aphorism model, for the observation being a known person named Joe:

$$\begin{bmatrix}\text{a measured quantity of Joe's health}\\ \text{a measured quantity of Bob's wealth}\\ \text{a measured quantity of Joe's wisdom}\end{bmatrix} = \text{model}\left(\begin{bmatrix}\text{a measured quantity of Joe's bedtime}\\ \text{a measured quantity of Joe's waketime}\end{bmatrix}\right)$$

Updated Definition of Prediction

Now we can hone our definition of prediction. For a

- new
- heretofore unseen
- future

observation, where the inputs have been measured / assessed but the output has not been measured / assessed,

$$\underbrace{\text{the guessed output measurements}}_{\text{prediction}} = \text{model}\left(\begin{array}{c}\text{the measured}\\ \text{inputs of an}\\ \text{observation}\end{array}\right)$$

$$\left[\begin{array}{c}\text{a guessed quantity of Bob's health}\\ \text{a guessed quantity of Bob's wealth}\\ \text{a guessed quantity Bob's wisdom}\end{array}\right] = \text{model}\left(\left[\begin{array}{c}\text{a measured quantity of bedtime}\\ \text{a measured quantity of waketime}\end{array}\right]\right)$$

Measurements as Variables

Instead of

> *Early to bed and early to rise make a man healthy, wealthy and wise.*

we have
Instead of "a measured quantity ..." we can use an algebraic *variable* (e.g. the symbol x) to denote the numerical quantity and now we have:

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = \text{model}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)$$

We will use the hat symbol (ˆ) to indicate a prediction.

Mathematical Model

Now that we have numbers and an equal sign, we have now created a *mathematical model* where the *model* now will be represented as a function, f where the inputs and outputs are specified implicitly.

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = f\left(x_1, x_2\right)$$