Session 20 – Prediction and the Many Model Thinker

C20.1 Prediction

Predictions:
(a) Made by individuals
(b) Made by collections of individuals

Wisdom of Crowds:
Categories: (R-squared metrics)
Linear Models: (data fitting)
Markov Models: (fixed state transition probability)

Quiz: What are three models we have discussed so far in this course that can help us make predictions? (a) Categories, (b) Linear Models, (c) Markov Models, and (d) Fermi's Method
Ans: (a), (b), (c)

Diversity Prediction Theorem:
Using a diversity of models is better for making predictions. It is an explanation of the 'wisdom of crowds'.

Quiz: Which one of the following concepts demonstrates how having many models allows us to make more accurate predictions than having only one model? (a) Categories, (b) Linear Models, (c) Markov Models, (d) Diversity Prediction Theorem
Ans: (d) Diversity Prediction Theorem

C20.2 Linear Models

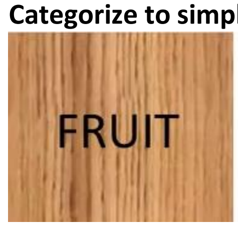
Predictive Models:
Categories: (R-squared metrics)
Linear Models: (data fitting)

Example: Predict height of Saturn rocket. People may typically try to compare with other 'tall' objects in their familiarity space.

'Lump to live' (categorize) metaphor

Categorization Method:
Recall: fruits and desserts
(Pear, 100), (Cake, 250), (Apple, 90), (Banana, 110),
(Pie, 350).
<u>**Recall Variation calculation**</u>: mean calories = 180
(Pear, $(100-180)^2 = 6400$), (Cake, $(250-180)^2 = 4900$), (Apple, $(90-180)^2 = 8100$), (Pie, $(350-180)^2 = 28,900$), (Banana, $(110-180)^2 = 4900$)
Total variation = sum = 53,200
Compare: Total variation = 53,200 vs ➔ ➔ ➔

Categorize to simplify:



Mean = 100
Variation = 200

Mean = 300
Variation = 5,000

Explained away 53,200 variation into 5,200.

R-Squared ➔ % of Variation
(use categorization to explain the observations to reduce variations and thus prediction).
R-squared = 1 – (sum of sorted variations)/(unsorted variation) = 1 – (5,200/53,200) = 90.2%

Linear Model Method:
Z = ax + by + cz + . . . , where Z is a dependent variable, & x, y, z, are independent variables

Calories in a Sub: (linear model)
Calories = bun + mayo + ham + chees + onion + tomato + lettuce.
(bun, 300), (mayo, 200), (ham, 50), (cheese, 100), (onion, 5), (tomato, 5), (lettuce, 5) ➔ total = 665

Quiz: 2 people are asked to estimate the population of New York City. The first person, Jonah, lives in Chicago, a densely populated city. He guesses 7.5 million. Nastia, the second person, lives in Bay City, Michigan, a relatively small town. She guesses 3 million. The actual population of New York City is 8,391,881. Which of the following may explain why Jonah's prediction was more accurate than Nastia's?

(a) To guess the population of New York City, Jonah and Nastia could have been using their own home cities as an analogy. Chicago, a large city (of population of about 2.9 million) is a better reference point than Bay City (of population about 2000).

(b) Nastia was probably only considering Manhattan when she was calculated the population of New York, while Jonah was considering the whole city. Therefore when Jonah linearly summed up his approximation for the populations of each burrow, he was probably closer to the right answer.

(c) Nastia probably categorized New York as 'mid-size city' whereas Jonah probably categorized it as 'very big city.'

(d) They both probably guessed randomly without reference. Jonah just happened to get lucky and pick a closer number.

Explanation: *Using appropriate analogies is important for accurately predicting things. It's natural to use one's hometown as an analogy when predicting the size of another city. This is much easier to do when the size of your city is closer to the size of the one you're predicting. For Jonah, his calculation could have been 'New York is probably just over 2.5x the size of where I live' and arrived at 7.5, which is still off, but closer than Nastia's prediction. Nastia, from Bay City, using her analogy had to consider 'New York is far larger than my town--maybe 1500 times the size?' and was farther off. Categories could have worked too, just not the example given. An appropriate example would have been: Jonah considers New York a 'very big city' and considers Chicago a 'big city'. Chicago has a population of 2.9 million and is 'big', so New York must have more than that to be considered 'very big'. However, Nastia might consider both Chicago and New York very big and therefore might estimate their populations to be closer to each other. B was wrong because there's no reason why Nastia should only consider Manhattan. That has nothing to do with the information given.*

Summary: These models predict on the basis of their variation from the observations: The Wisdom of Crowds depends upon the variations in the models (categorizations, linear, percolation, Markov, etc.) held by each individual. So one metric is the variations within the crowd.

C20.3 Diversity Prediction Theorem

Collective Wisdom: by way of the **Diversity Prediction Theorem:** ***Observation:*** (a) more accurate individuals imply more accurate predictions, (b) more diversity in crowd implies more accurate predictions. Question: if Crowd Accuracy (CA) = Individual Accuracy(IA) + Crowd Diversity (CD) or CA = IA + CD What matters more?	**Example:** Prediction of number of diners coming today – (Amy, 10), (Belle, 16), (Carlos, 25). **Crowd Average:** $\boldsymbol{c = (1/n)\sum S_i = 17}$, And given **actual value (θ) = 18** **Average Individual Error:** **$(1/n)\sum(S_i - \theta)^2$** (Amy, $(10\text{-}18)^2$=64), (Belle, $(16\text{-}18)^2$=4), (Amy, $(25\text{-}18)^2$=49) for a $\sum R^2$ = 117 and average R^2, $\sum R^2/n$ = 39; aka $(1/n)\sum(S_i - \theta)^2$ below **Crowd Error:** **$(c - \theta)^2$** (Crowd, $(17\text{-}18)^2$=1), so R^2_{crowd} (1) < $R^2_{best\ individual}$ (4)
In Example: Look at Diversity (Variation) Measure Crowd Diversity, $\sum R^2/n$, using the crowd average, c, that is, $(1/n)\sum(S_i - c)^2$	**Crowd Diversity:** **$(1/n)\sum(S_i - c)^2$** (Amy, $(10\text{-}17)^2$=49), (Belle, $(16\text{-}17)^2$=1), (Amy, $(25\text{-}17)^2$=64) for a R^2 sum = 114, & $\sum R^2/n$ = 38 Diversity = $\sum R^2/n$ aka $(1/n)\sum(S_i - c)^2$ below.

Experiment Observation: Crowd's Error = 1, Average error (individual) = 1, Diversity = 38 AND IN THIS CASE
1 = 39 -38 BUT THIS IS ALWAYS TRUE!
Diversity Prediction Theorem:
Crowd's error = Average Error – Diversity

c = Crowd Average, θ is the true value, so $(c-\theta)^2$ is the crowd's squared error. S_i is individual i *'s* prediction

Crowd Error:
crowd squared error =
average individual squared error – crowd diversity

$$(c-\theta)^2 = \frac{1}{n}\sum_{i=1}^{n}(S_i-\theta)^2 - \frac{1}{n}\sum_{i=1}^{n}(S_i-c)^2$$

$\theta = true\ value$
$S_i = Individual\ i's\ prediction$
$c = Crowd's\ prediction$ (average of individual predictions, $c = (1/n)\sum S_i$)

Quiz: Anisha, Matt and Wendy are on a game show. They are trying to guess the price of a new car. Anisha's prediction is \$50,000. Matt's prediction is \$30,000. Wendy's prediction is \$43,000. The actual price is \$41,050. They win the car if their crowd squared error is less than 5,000. Do they win the car? What is the crowd diversity?

(a) No, they don't win the car. Crowd Diversity= 41,000
(b) Yes, they win the car. Crowd Diversity=68,666,666.67
(c) No, they don't win the car. Crowd Diversity = 45,678,439.34
(d) Yes, they win the car. Crowd Diversity = 41,000

Analysis: looking for $(c - \theta)^2$. θ = \$41,050, c = (1/3)(50,000 + 30,000 + 43,000) = 41,000, i.e., c = 41,000 ➔
Crowd (squared) Error = $(41{,}000-41{,}050)^2 = (50)^2 = 2500$, which is less than 5,000 AND
Crowd diversity = $(1/3)\{(50{,}000 – 41{,}000)^2 + (30{,}000 – 41{,}000)^2 + (43{,}000 – 41{,}000)^2\}$ = 68,666,666.67
Ans: (b)
Explanation: *they win the car if their crowd error is less than 5000. Their crowd squared error is: (average-actual)^2: $[(30{,}000+50{,}000+43{,}000)/3)-41{,}050]^2 = 50^2 = 2500$. Crowd diversity = $(1/3)\{(50{,}000 – 41{,}000)^2 + (30{,}000 – 41{,}000)^2 + (43{,}000 – 41{,}000)^2\}$ = 68,666,666.67*

UK steer weight challenge: Crowd guess of 1100 pound steer was only off by 1 lb.
Crowd's error = Average Error – Diversity
Actual Formula yields

0.6 = 2,956.0 – 2,955.4

BUT NOTE these are squared errors (R-squared)

Crowd Error = Average Error - Diversity

CE = AE – Div where diversity matters a lot

To be a 'Wisdom of Crowds' phenomena,
CE has to be small. Also AE has to be large, otherwise there is no surprise wisdom. Therefore Diversity must be large.

$$\boldsymbol{CE = AE - Div}$$
$$Small = Large - Large$$

Above example at the left shows this relationship.

How do you get large diversity is by using many different models?

Wisdom of crowds comes from reasonably smart people using a diversity of models.

Both contribute equally to the Wisdom of Crowds.

Quiz: If diversity increases, what must happen to the average error in order to retain the same crowd error? (a) The average error must decrease. (b) The average error must increase as well.
Ans: (b) The average error must increase as well.
Explanation: *Remember that Crowd Error = Average Error - Diversity.*
So if diversity increases, but we want the result (crowd error) to remain the same, then the average error must increase along with the diversity.
If the average error does not increase, on the other hand, but diversity still does, then the crowd error would decrease. [See 20.3, "Diversity Prediction Theorem"]

Quiz: A more diverse crowd creates a more accurate crowd prediction. (a) True, (b) False
Ans: (a) True

Madness of Crowds: If CE is to be large, need large Average Error and small Diversity. Small Diversity from like-minded people who are all wrong.

$$\boldsymbol{CE = AE - Div}$$
$$Large = Large - Small$$

C20.4 The Many Model Thinker

Why become a many model thinker? ***Reason #1*** – Intelligent Citizen of the World.	**Growth models –** just invest in capital and add innovation. Innovation has a squaring effect. **Colonel Blotto** – How in some contexts it makes sense to add new dimensions.
Reason # 2 – Becoming Clear Better Thinkers Markov Models – trends not always linear. History doesn't always matter. Intervention in state can be ineffective whereas intervention in probabilities is the driver.	**Tipping Points** – Don't confuse growth (slopes) with tips **Path Dependence** – on how things unfold. **Chaotic Models** **Percolation Models**: **SIS/SIR Models:**

Quiz: Which of the following models help us to become clearer, better thinkers? Choose all that apply. (a) Markov Models, (b) Tipping Points, (c) Path Dependence, (d) Percolation Models
Ans: (a), (b), (c), (d)

Reason # 3 – Understand and Use Data **Categorization** **Linear Model** **Growth Model** **Prediction** to get Wisdom of Crowds **Markov models** also apply	**Quiz**: Which of the following models help us to understand and use data? Choose all that apply. (a) Categorical Models, (b) Linear Models, (c) Prediction Models, (d) Markov Models **Ans:** (a), (b), (c), (d)
Reason # 4 – Decide, Strategize, and Design **Decision Theory Models** **Game Theory Models** - Prisoner's Dilemma **Collective Action** **Mechanism Design** – to help design institutions, contracts, & Policy, Truth revelation, or incentives for desired actions (possible or not possible) **Auctions** – types and results (rational actors indifferent)	**How do People Behave? *(a) Rational, (b) Follow Rules, (c) Psychological Biases*** In some cases, behavioral assumptions have a huge effect on the model. In other cases like exchange markets, didn't matter at all – generally same outcome in way people behave given reasonable coherence in actions they take. **Quiz:** What are the three models for how people behave? (a) Rational, (b) Behavioral, (c) Rule-based, (d) Race to the bottom **Ans:** (a), (b), (c)

Thinking models often aggregate differently than we expect. When we aggregate, can get different interesting things. Recall Wolframs and four classes of systems (static (go to equilibrium), cyclic, random, complex) Learned how different systems can produce different outcomes. How to intervene. What actions people are likely to take. How events are likely to unfold. How do we collectively understand what's going to happen in different parts of the world?
Quiz: Which of the following is a potential outcome of a system? Choose all that apply. (a) Chaos, (b) Equilibria, (c) Patterns, (d) Randomness **Ans:** (a), (b), (c), (d) **Note:** (a) randomness is the same as (d) chaos as used in this course.
Summary: By constructing these models they're this crutch for us, they help us clarify our thinking, we're better thinkers, we use data better, we can design and strategize better and we're just more intelligent citizens.