- Tap the + button
- Write the first word that comes to mind.
- Use your phone’s predictive text feature to generate 10 to 20 more words.

Math & Prediction

Weather, Language Models & More

11/7/2023

Weather Forecasting

- Weather is an unpredictable event. We don't fully know tomorrow's weather, next week weather, or next month weather.
- Suppose that all the information we have access to is of **past weather data**
- **Goal:** predict whether the next few days are going to be **cold** or **warm**
    - **Level 1:** count cold vs warm days and compute probability of each
    - **Level 2:** use info about today's weather. If it is cold, use past data to see what is more likely after a cold day

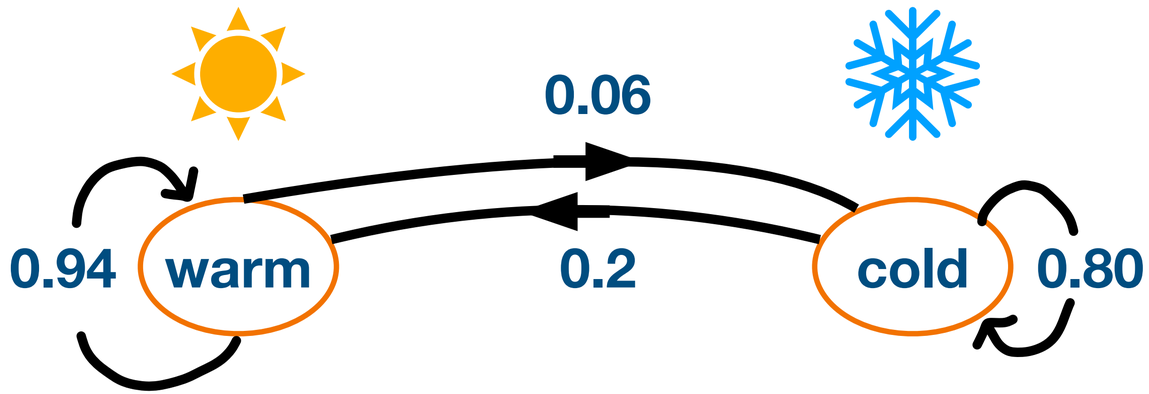
The Data

- Collect the daily data, for example of the year 2023 up until today (309 days).
- There where 238 warm days (0.77%)
- There where 71 cold days (0.33%)
- There is a probability of 0.77 of it being warm, and 0.33 of being cold

The Data

- Of the 71 cold days, 57 were followed by a cold day (80%)
- Of the 238 warm days, 223 were followed by a warm day (94%)

Note: in New England warm means above 5 Celsius / 40 Fahrenheit



Building the Transition Matrix

- If today is cold, then tomorrow there is a 80% of it being cold again, and 20% warm
- If today is warm, then tomorrow there is a 94% of it being warm again, and 6% cold
- Transition matrix of **probabilities**:

Today \ Tomorrow	**Cold**	**Warm**
Cold	80 %	20 %
Warm	6 %	94 %

Weather Upcoming Days

- Today is warm. Tomorrow there is a 94% of it being warm again, and 0.06% cold
- The day after tomorrow:
    - Chance of being warm: $0.94 \times 0.94 + 0.06 \times 0.2 = 0.8956$
    - Chance of being cold: $0.94 \times 0.06 + 0.06 \times 0.8 = 0.1044$
- Transition matrix of **probabilities**:

Day After Tomorrow

Today

	Cold	**Warm**
Cold	65,2 %	34,8 %
Warm	10,44 %	89,56 %

Weather Forecast

More accurate:

- More variables: Cold/Warm, Dry/Humid, Sunny/Cloudy, Windy/Calm
- Seasonal variability
- Data from previous years
- Other factors…

Language Models

Language Models

- Wikipedia: “ A **language model** is a model of a natural language[1] that can **generate probabilities** of a series of words, based on the **text samples** in one or multiple languages it was trained on.”

- Mathematics: Probability/Statistics, Linear Algebra, Graph Theory,…

Language Models

- **Statistical/Probabilistic:** Given a word or a sentence, find what word has the highest probability of coming next.

- **Recurrent Neural Networks:** Uses word embeddings to transform words into vectors.

- **Large Language Models:** Notable examples include OpenAI's GPT models (e.g., GPT-3.5 and GPT-4, used in ChatGPT), Google's PaLM (used in Bard), and Meta's LLaMa, as well as BLOOM, Ernie 3.0 Titan, and Anthropic's Claude 2.

Can we learn a new language in 15 minutes?

Wiele dziedzin nauki i technologii w wieku

- Three types of characters:

  blanks (**B**), consonants (**C**) and vowels (**V**) (a, e, i, o, u, y)

- Your task: Simulate 10 **more characters** using the last one (u) as your starting point:

… wiekuBCVCCVBCVC

  - Find the transition matrix.
  - Use dice to make a random choice (according to the probability).