CS-131 - Bayesian Classifier – Handout

Introduction

This handout explains the step-by-step process and some important concepts of a Bayesian classifier, particularly for sequential data classification. It includes a running example scenario to help develop intuition and answer common questions that you might have.

Important points and definitions

1. **Naïve Bayesian Classification:** Uses Bayes' theorem to calculate the probability of an observation belonging to a specific class based on **prior knowledge** and **current evidence**.
2. **Recursive Estimation:** The posterior probability from the previous step becomes the prior for the next step, allowing the classifier to account for **temporal** dependencies.
3. **Likelihoods:** These are probabilities of observing a specific feature value given a class. For instance, the likelihood of a specific velocity being associated with a bird versus an airplane.
4. **Transition Probabilities:** They reflect the likelihood of staying in the same class versus transitioning to another class over time. This adds temporal consistency to the classifier.
5. **Normalization:** Ensures probabilities are valid (sum to 1) after each update.
6. **Handling Missing Data (NaN):** Missing values in the dataset should be ignored or handled gracefully without disrupting the calculation process.

Running example

Scenario: Classify a sequence of temperatures as either Hot (H) or Cold (C) based on likelihoods and transition probabilities.

Setup:

- **Likelihoods:**

$$P(\text{Temp} \mid H) : \text{Probability of observing a temperature given it is hot.}$$

$$P(\text{Temp} \mid C) : \text{Probability of observing a temperature given it is cold.}$$

- **Transition Probabilities:**

$$P(H_{t+1} \mid H_t) = 0.9, \quad P(C_{t+1} \mid C_t) = 0.9$$

$$P(H_{t+1} \mid C_t) = 0.1, \quad P(C_{t+1} \mid H_t) = 0.1$$

- **Priors:** Initial probabilities are set to:

$$P(H) = 0.5, \quad P(C) = 0.5$$

- **Data:** Sequence of temperatures: [55, 65, 75, NaN, 85]

Dry run for the algorithm

Iteration 1 (Temperature = 55°F)

1. **Observation Likelihoods:**

$$P(\text{Temp} = 55 \mid H) = 0.1, \quad P(\text{Temp} = 55 \mid C) = 0.8$$

2. **Prior Probabilities:**

$$P(H) = 0.5, \quad P(C) = 0.5$$

3. **Posterior Update:**

$$P(H \cap \text{Temp}) = P(\text{Temp} \mid H) \cdot P(H) = 0.1 \cdot 0.5 = 0.05$$
$$P(C \cap \text{Temp}) = P(\text{Temp} \mid C) \cdot P(C) = 0.8 \cdot 0.5 = 0.4$$

4. **Normalization:**

$$P(H \mid \text{Temp}) = \frac{0.05}{0.05 + 0.4} = 0.11$$
$$P(C \mid \text{Temp}) = \frac{0.4}{0.05 + 0.4} = 0.89$$

Iteration 2 (Temperature = 65°F)

1. **Observation Likelihoods:**

$$P(\text{Temp} = 65 \mid H) = 0.3, \quad P(\text{Temp} = 65 \mid C) = 0.7$$

2. **Transition Probabilities:** Update priors using the transition probabilities:

$$P(H_t) = P(H_{\text{prev}}) \cdot 0.9 + P(C_{\text{prev}}) \cdot 0.1$$
$$P(C_t) = P(C_{\text{prev}}) \cdot 0.9 + P(H_{\text{prev}}) \cdot 0.1$$

3. **Posterior Update:**

$$P(H \cap \text{Temp}) = P(\text{Temp} \mid H) \cdot P(H_t)$$
$$P(C \cap \text{Temp}) = P(\text{Temp} \mid C) \cdot P(C_t)$$

4. **Normalization:** Normalize the probabilities to ensure they sum to 1.

Iteration 3 (Temperature = 75°F)

Repeat the same steps: compute observation likelihoods, update priors using transitions, calculate posteriors, and normalize.

Handling Missing Data (NaN)

When a value is missing (e.g., NaN), skip the step and retain the prior probabilities from the previous iteration.

Iteration 5 (Temperature = 85°F)

Use the final observation likelihoods and the updated priors to compute the final posterior probabilities.

Final Classification

- The posterior probabilities evolve at each step based on observations and transitions.
- The classification for each time step is determined by the higher posterior probability.
- Missing data is skipped without disrupting the recursive process.

Summary

This step-by-step process highlights how to implement a Bayesian classifier for sequential data, including transitions and handling missing data. The final classification depends on maximizing posterior probabilities at each time step.

So what did we do..

1. Use likelihoods to measure the compatibility of observations with each class.
2. Incorporate transitions to account for temporal consistency.
3. Normalize probabilities to ensure they remain valid.
4. Handle missing data gracefully to maintain the integrity of the classifier.

You should apply this understanding to the given radar trace classification task, making sure a logical approach to additional feature extraction and likelihood integration.

Good luck!