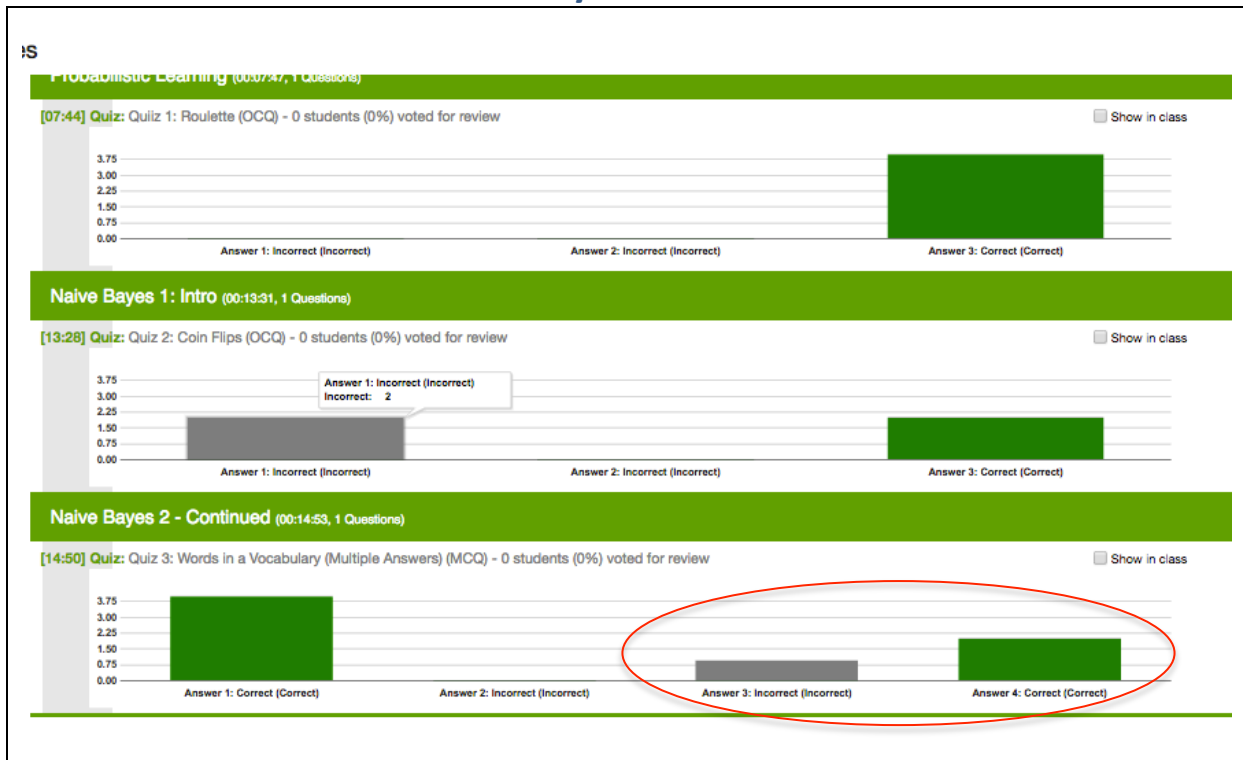


## Feedback: Lecture 6 – Naïve Bayes



OBS: one student have missed one possible correct answer.

### Quiz



- ▶ The pockets of a roulette wheel are numbered 0–36 with 18 red numbers and 18 black numbers.
- ▶ Which prediction will have the lowest error rate over a long sequence of (fair) spins?
  - The next number is 24
  - The next number is black
  - The next number is not 24

First quiz: Unproblematic.

- Answer 1: Incorrect: This prediction has probability  $1/37$  and will on average have an error rate around 98%.
- Answer 2: Incorrect: This prediction has probability  $18/37$  and will on average have an error rate around 51%.
- Answer 3: Correct: This prediction has probability  $36/37$  and will on average have an error rate around 2%.

## Quiz

- ▶ Do you remember the new sequence of coin flips: 001
- ▶ Here are the probabilities:  
 $P(C_1) = 4/7$   $P(C_2) = 3/7$   $P(X_i = 1|C_1) = 12/16$   $P(X_i = 1|C_2) = 6/18$
- ▶ Which coin is most probable?
  - $P(C_1|X = 001) > P(C_2|X = 001)$
  - $P(C_1|X = 001) = P(C_2|X = 001)$
  - $P(C_1|X = 001) < P(C_2|X = 001)$

Answer 1: Incorrect.

Answer 2: Incorrect.

**Answer 3: Correct: The class probabilities are roughly equal ( $4/7 \approx 3/7$ ), but the probability of getting 001 with C1 is only  $0.25 \times 0.25 \times 0.75$ , compared to  $0.67 \times 0.67 \times 0.33$  for C2, so C2 is more probable.**

## Quiz

- ▶ Suppose our entire vocabulary consists of four words:  
**buy, sell, drug, multinomial**
- ▶ Suppose we have the following frequency counts in our collection of spam documents:  
 $n_{\text{buy}} = 50$   $n_{\text{sell}} = 10$   $n_{\text{drug}} = 40$   $n_{\text{multinomial}} = 0$
- ▶ Which of the following statements are correct?
  - The MLE of  $P(\text{multinomial}|\text{SPAM}) = 0$
  - The MLE of  $P(\text{multinomial}|\text{SPAM}) = 1$
  - The MAP of  $P(\text{multinomial}|\text{SPAM})$  with Laplace smoothing =  $1 / 100$
  - The MAP of  $P(\text{multinomial}|\text{SPAM})$  with Laplace smoothing =  $1 / 104$

1 **Correct:** The MLE is  $0 / 100 = 0$ .

2 Incorrect.

3 Incorrect.

4 **Correct:** The MAP is  $(0+1)/(100+4)$ .