problem 2.

1) 3-grams: $p_r(W \mid W_{i-1}, W_{i-2}) = \frac{\text{number of sequence } (W_{i-2}, W_{i-1}, W)}{\text{number of sequence } (W_{i-2}, W_{i-1})}$. (MLE).

clearly. it's follows Multinomal Distibution.

2) ① For a given data set, we first need to ~~split~~ devided it into training and testing data. randomly.

② For the training data set: ~~we~~ calculate the ~~portion~~ porportion of email and proportion of span: $P(\text{email})$ and $p(\text{spam})$

And $P(\text{Spam}) + P(\text{email}) = 1$

③ For the new data set we make the decision based on Lazy learning Approach and idea of bayes theorem: $\hat{y} = \cancel{\text{argmax}} \underset{y}{\text{argmax}}\, p(y \mid \text{new data}) = \underset{y}{\text{argmax}}\, p(y)\, p(\text{new data} \mid y)$.

$$p(\text{spam} \mid \text{new data}) \propto p(\text{spam}) \times p(\text{new data} \mid \text{spam})$$
$$= p(\text{spam}) \times \prod_{i=3}^{m} (\text{3-gram probability of } i)$$
$$= p(\text{spam}) \times \prod_{i=3}^{m} \frac{\#\text{ of sequence } (W_{i-2}, W_{i-1}, W)}{\#\text{ of sequence } (W_{i-2}, W_{i-1})}$$

~~W~~ W is the words in the new data set.

in the same way: $p(\text{email} \mid \text{new data}) \propto p(\text{email}) \times p(\text{new data} \mid \text{email})$

$$= p(\text{email}) \times \prod_{i=3}^{m} \frac{\#\text{ of sequence } (W_{i-2}, W_{i-1}, W)}{\#\text{ of sequence } (W_{i-2}, W_{i-1})}$$

④ Finally, $y = \begin{cases} \text{span} & \text{if } p(\text{spam} \mid \text{new data}) \geq p(\text{email} \mid \text{new data}) \\ \text{email} & \text{otherwise.} \end{cases}$

3) As mentioned above. for a new message. $\hat{y} = \underset{y}{\text{argmax}}\, p(y \mid \text{new data}) = \underset{y}{\text{argmax}}\, p(y) \cdot p(\text{new data} \mid y)$

$y = \begin{cases} \text{span} & \text{if } p(\text{spam} \mid \text{new data}) \geq p(\text{email} \mid \text{new data}) \\ \text{email} & \text{otherwise} \end{cases}$

By the way, we can use this method for test data and calculate the Accuracy of the model.