Name: _____   SID: _____

COMP 4446 / 5046 Quiz 1 (week 4) - version A

Please read the instructions on the screen before beginning. Select **all** correct options by filling in circles: ●. If you make a mistake, draw an X over the circle: ✖.

1. (1 mark) What benefits does a One-Hot Vector have over a Sparse Vector?
   - ○ Lower memory usage
   - ○ Faster to calculate similarity
   - ◐ Faster to modify a value
   - ◐ Faster to check a value

2. (1 mark) Which of the following is true of a Model?
   - ✓ It calculates the score of an (input, output) pair
   - ○ It finds the correct output
   - ✗ It calculates the score for an input
   - ○ It finds a high scoring output

3. (1 mark) Select all of the statements that are true about recurrent neural networks:
   - ○ They can only predict one output for each input
   - ○ They do not suffer from the vanishing gradient problem
   - ◐ They can easily and effectively handle inputs of different lengths
   - ◐ They contain a non-linearity

4. (1 mark) For each scenario below, you are deciding what metric should be optimised to keep users happy. Of the options provided, which is best? Note that the rubric for this question will consider both responses together (ie., it will not be 0.5 each).

   Identifying documents that are relevant to a trial. Here, a true positive is a relevant document that is labelled as relevant. It is important that no relevant documents are missed.

   ○ Precision   ◐ Recall   ○ F-Score   ○ Accuracy

   Classifying whether a website is written in Italian for a study of how commonly Italian is used online. Here, a true positive is an Italian website that is labelled as Italian. It is important that the results are representative of the web.

   ✗ Precision   ○ Recall   ✓ F-Score   ○ Accuracy

5. (1 mark) In workshop 2, task 2, we calculated how often people lied. Using the lines below, implement code to take a list of labels and calculate what fraction are lies.

```
1  fraction = total / count          6  if label == "lie": total += 1
2  fraction = count / total          7  if label == "lie": count += 1
3  total += 1                        8  if label == "truth": total += 1
4  count += 1                        9  if label == "truth": count += 1
5  total, count = 0, 0              10  for label in labels:
```

   In each row below, choose one line of code by filling in the appropriate circle. You do not have to use all the given code and you do not have to use all rows below.

1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○
○	○	○	○	○	○	○	○	○	○
○	○	○	○	○	○	○	○	○	○
○	○	○	○	○	○	○	○	○	○
○	○	○	○	○	○	○	○	○	○
○	○	○	○	○	○	○	○	○	○

   (handwritten: 5, 10, 3, 7, 2)

COMP 4446 / 5046 Quiz 1 (week 4) - version B

Please read the instructions on the screen before beginning. Select **all** correct options by filling in circles: ●. If you make a mistake, draw an X over the circle: ✖ .

1. (1 mark) What advantages does a Map have over a One-Hot Vector?
   - ○ Faster to check a value
   - ◉ Faster to calculate similarity
   - ○ Faster to modify a value
   - ◉ Lower memory usage

2. (1 mark) Which of the following is true of an Inference Method?
   - ○ It calculates the score for an input
   - ✓ It finds a high scoring output
   - ○ It calculates the score of an (input, output) pair
   - ✗ It finds the correct output

3. (1 mark) Select all of the statements that are true about feedforward networks / multi-layer perceptrons:
   - ◉ They contain a non-linearity
   - ○ They do not suffer from the vanishing gradient problem
   - ○ They can only predict one output for each input
   - ○ They can easily and effectively handle inputs of different lengths

4. (1 mark) For each scenario below, you are deciding what metric should be optimised to keep users happy. Of the options provided, which is best? Note that the rubric for this question will consider both responses together (ie., it will not be 0.5 each).

   Spam detection for a client who does not want to miss any real mail. Here, a true positive is a message that was correctly labelled as spam. It is important that no real mail is labeled as spam.
   - ◉ Precision   ○ Recall   ○ F-Score   ○ Accuracy

   Filtering applicants for the cast of a play where the director wants to save time but still form the best group. Here, a true positive is a good applicant that was correctly included in the list to consider. It is important that all good applicants are seen.
   - ○ Precision   ◉ Recall   ○ F-Score   ○ Accuracy

5. (1 mark) In workshop 2, task 2, we calculated how often people lied. Using the lines below, implement code to take a list of labels and calculate what fraction are the truth.

```
1  fraction = count / total          6  if label == "truth": total += 1
2  fraction = total / count          7  total += 1
3  if label == "lie": total += 1     8  count += 1
4  if label == "lie": count += 1     9  for label in labels:
5  if label == "truth": count += 1  10  total, count = 0, 0
```

   To give your answer, fill in ONE circle in each row below. You do not have to use all the given code and you do not have to use all rows.

1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○
○	○	○	○	○	○	○	○	○	○
○	○	○	○	○	○	○	○	○	○
○	○	○	○	○	○	○	○	○	○
○	○	○	○	○	○	○	○	○	○
○	○	○	○	○	○	○	○	○	○

COMP 4446 / 5046 Quiz 2 (week 6) - version A

If the choices have ◯ then select exactly one option. If the choices have ☐, select all correct options. Indicate your answer by filling the shape, e.g., ●. If you make a mistake, draw an X over your answer, e.g., ✖ .

1. (1 mark) Which of the following are true of top-1 and top-K sampling? Select all true statements.
   - ☐ They both use randomness to choose the output.
   - ☐ Their filtering step is the same, but they sample differently.
   - ☐ Neither one considers the probability distirbution when filtering.
   - ☐ They are both greedy methods.

2. (1 mark) Which of the following are true of Keras? Select all true statements.
   - ☐ It is built on top of Tensorflow.
   - ☐ It is a library for neural networks.
   - ☐ It requires a GPU to run code.
   - ☐ It supports training and applying models.

3. (1 mark) Consider the sentence, "Joe located people with a dog." If we have a parse with an edge from "with a dog" to "located", what does that mean? Select one option.
   - ◯ Joe used a dog to locate the people.
   - ◯ Joe and the people have dogs.
   - ◯ The people have a dog.
   - ◯ None of the above options are correct.

4. (1 mark) Which of the following are true of the Viterbi algorithm? Select all true statements.
   - ☐ If we modify it to track the previous two labels, the time complexity is $O(|words| * |labels|^3)$
   - ☐ As each word is processed, the algorithm can tell you what the label should be for that word.
   - ☐ At each step, for each possible label, we store a previous label and probability.
   - ☐ It can be used with RNNs, but not feed forward networks.

5. (1 mark) Using the lines below, implement one step of the Viterbi algorithm. `tok` is the current token. `pos` is the current position in the sequence. `result` stores the algorithm's output. Provide your answer by filling in one circle in each row of the grid below.

```
1  for label in labels:              9  best = 0
2  result[pos][label] = best        10  score *= result[pos-1][0]
3  score = 1                        11  score *= result[pos-1][plabel][0]
4  score *= tmodel(label)           12  if score > best[0]:
5  score *= tmodel(plabel, label)   13  for plabel in labels:
6  best = (score, plabel)           14  score *= emodel(label)
7  best = score                     15  score *= emodel(tok)
8  best = (0, None)                 16  score *= emodel(tok, label)
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

COMP 4446 / 5046 Quiz 2 (week 6) - version B

If the choices have ○ then select exactly one option. If the choices have □, select all correct options. Indicate your answer by filling the shape, e.g., ●. If you make a mistake, draw an X over your answer, e.g., ✖ .

1. (1 mark) Which of the following are true of top-P and top-K sampling? Select all true statements.

   ☐ Neither one considers the probability distirbution when filtering.
   ☐ They are both greedy methods.
   ☐ They both use randomness to choose the output.
   ☐ Their filtering step is the same, but they sample differently.

2. (1 mark) Which of the following are true of PyTorch? Select all true statements.

   ☐ It is a library for neural networks.
   ☐ It supports training and applying models.
   ☐ It requires a GPU to run code.
   ☐ It is built on top of Tensorflow.

3. (1 mark) Consider the sentence, "Zach found people with a telescope." If we have a parse with an edge from "with a telescope" to "found", what does that mean? Select one option.

   ○ Zach and the people have a telescope.
   ○ Zach was using a telescope.
   ○ The people have a telescope.
   ○ None of the above options are correct.

4. (1 mark) Which of the following are true of the Viterbi algorithm? Select all true statements.

   ☐ It is not compatible with models that consider the entire output sequence at once.
   ☐ It cannot be used with a neural network model.
   ☐ The standard version has a time complexity of $O(|words| * |labels|^2)$
   ☐ When the algorithm is halfway through, it can tell you what the first half of the output will be.

5. (1 mark) Using the lines below, implement one step of Beam search. `beam` contains the current beam. `k` is the intended beam size. `token` is the current token. Provide your answer by filling in one circle in each row of the grid below.

```
1  for label in labels:            9  beam = new_beam
2  new_beam.sort(reverse=True)    10  beam = new_beam[:k]
3  new_beam.sort()                11  for option in labels:   alt
4  option = item + [label]   alt  12  for item in beam:
5  new_beam = []                  13  new_beam.append((score, option))
6  score = get_score(option)      14  new_beam.append(score)
7  score = get_score(token)       15  new_beam.append(option)
8  score = get_score(token, option)
```

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Name:		SID:	

COMP 4446 / 5046 Quiz 3 (week 8) - version A

If the choices have ○ then select exactly one option. If the choices have □, select all correct options. Indicate your answer by filling the shape, e.g., ●. If you make a mistake, draw an X over your answer, e.g., ✖.

1. (1 mark) How are self-attention and cross-attention the same? Select all true statements.
   - ☐ The way the output is used.
   - ☐ The way the weights (α) are used.
   - ☐ The source of the keys, values, and queries.
   - ☐ The similarity calculation.

2. (1 mark) What is the purpose of RoPE? Select all true statements.
   - ☐ To introduce a non-linear step in the self-attention process.
   - ☐ To give flexibility in which parts of the vectors are used in each step.
   - ☐ To make learning easier by making it easier for gradients to propagate.
   - ☐ To enable self-attention to account for the position of a word in the input.

3. (1 mark) Why is the output of the encoder passed into every step of the decoder? Select all true statements.
   - ☐ To enable parallel processing in the decoder.
   - ☐ To make the dimensionality of inputs to the cell match at all steps.
   - ☐ To help the model not produce output that diverges too much from the original input.
   - ☐ To help the model not produce output that is completely different from the true answer.
   - ☐ To improve the speed of training.

4. (1 mark) Which of these parts of the transformer help make training smoother? Select all true statements.
   - ☐ Residual connections
   - ☐ Layer normalisation
   - ☐ Positional encoding
   - ☐ Feedforward layers
   - ☐ Self-attention

5. (1 mark) Using the lines below, implement code that uses spaCy to count occurrences of organisations in a provided string `text`. Provide your answer by writing the line numbers in the boxes to the right, in order, top to bottom.

```
1  for token in doc:
2  for ent in doc.ents:
3  doc = nlp.entities(text)
4  doc = spaCy(text)
5  doc = nlp(text)
6  nlp = spacy.load("en_core_web_sm")
7  counts = {}
8  if ent.label_ == "ORG":
9  if token.ent == "ORG":
10 counts[token] = countes.get(token, 0) + 1
11 counts[ent.text] = countes.get(ent.text, 0) + 1
```


COMP 4446 / 5046 Quiz 3 (week 8) - version B

If the choices have ○ then select exactly one option. If the choices have □, select all correct options. Indicate your answer by filling the shape, e.g., ●. If you make a mistake, draw an X over your answer, e.g., ✖.

1. (1 mark) How do self-attention and cross-attention differ? Select all true statements.
   - □ The similarity calculation.
   - ☑ The way the output is used.
   - □ The way the weights (α) are used.
   - ☑ The source of the keys, values, and queries.

2. (1 mark) What is the purpose of Q, K, and V matrices? Select all true statements.
   - ☒ To enable self-attention to account for the position of a word in the input.
   - ☑ To give flexibility in which parts of the vectors are used in each step.
   - □ To introduce a non-linear step in the self-attention process.
   - □ To make learning easier by making it easier for gradients to propagate.

3. (1 mark) What is the purpose of teacher forcing? Select all true statements.
   - ☑ To help the model not produce output that is completely different from the true answer.
   - ☑ To improve the speed of training.
   - □ To enable parallel processing in the decoder.
   - □ To make the dimensionality of inputs to the cell match at all steps.
   - □ To help the model not produce output that diverges too much from the original input.

4. (1 mark) Which of these parts of the transformer always contain a weight matrix that is learned in training? Select all true statements.
   - □ Positional encoding
   - ☑ Feedforward layers
   - ☒ Layer normalisation
   - ☑ Self-attention
   - □ Residual connections

5. (1 mark) Using the lines below, implement code that uses spaCy to count occurrences of pronouns in a provided string `text`. Provide your answer by writing the line numbers in the boxes to the right, in order, top to bottom.

```
1  nlp = spacy.load("en_core_web_sm")
2  counts[token.text] = counts.get(token.text, 0) + 1
3  counts[token.pos_] = counts.get(token.pos_, 0) + 1
4  counts = {}
5  for token in spaCy(text):
6  for pronoun in nlp(text):
7  for token in nlp(text):
8  for pronoun in spaCy(text):
9  if token.pos_ == "PRON":
10 if token.text in ["I", "we", "he", "she", "it", "you"]:
```

COMP 4446 / 5046 Quiz 4 (week 10) - version A

Please see the screen for your questions. Note, we have intentionally provided boxes that are larger than necessary. Your answer does not need to fill the whole box.

1. (1 mark) When providing examples for in-context learning, what way of ordering them would hurt performance?

2. (1 mark) When does LoRA improve performance and on what measure?

3. (1 mark) What is a key difference between pre-training and training?

4. (1 mark) When querying the Pinecone database, how is the query represented?

Name:

SID:

COMP 4446 / 5046 Quiz 4 (week 10) - version B

Please see the screen for your questions. Note, we have intentionally provided boxes that are larger than necessary. Your answer does not need to fill the whole box.

1. (1 mark) When providing examples for in-context learning, what modification to the labels would only slightly impact performance?

2. (1 mark) When does LLM.int8 improve performance and on what measure?

3. (1 mark) What is a key similarity between pre-training and training?

4. (1 mark) In a RAG system, how are the retrieved documents used during response generation?

Name:		SID:	

COMP 4446 / 5046 Quiz 5 (week 12) - version A

Please see the screen for your questions. Note, we have intentionally provided boxes that are larger than necessary. Your answer does not need to fill the whole box.

1. (1 mark) What is the purpose of instruction tuning for large language models?

2. (1 mark) Why do we usually collect 2+ labels per example for the test set?

3. (1 mark) What advantage does Cohen's Kappa have over just counting what fraction of labels match?

4. (1 mark) Where does the data for preference optimisation come from?

COMP 4446 / 5046 Quiz 5 (week 12) - version B

Please see the screen for your questions. Note, we have intentionally provided boxes that are larger than necessary. Your answer does not need to fill the whole box.

1. (1 mark) How does instruction tuning differ from regular fine-tuning?

2. (1 mark) Why do we usually collect only 1 label per example for the training set?

3. (1 mark) If a dataset has a large label imbalance, what impact does that have on Cohen's Kappa?

4. (1 mark) Where does the data for instruction tuning come from?