1. (1 point) Which of the following is a valid term-weighting formula used in vector-space models:

   A. $f_{i,j}$
   B. $f_{i,j} \times \log(1 + N/df_i)$
   C. $(1 + \log f_{i,j}) \times \log(N/df_i)$
   D. $\log f_{i,j} \times \log(1 + df_i)$
   E. None of the above.

1. B C

2. (1 point) What are the advantages of tokenizers like BPE Tokenizer (or Wordpiece tokenizer) in comparison to traditional tokenizers using delimiters like space,tab,hyphens etc.:

   A. Handling common subwords so that they can reduce the need for stemming.
   B. Efficient tokenization of text content in large corpus.
   C. Much smaller vocabulary size.
   D. Data-centric tokenization (such as BPE/Wordpiece) help handle wider variety of languages such as German, Hindi, Chinese etc.
   E. It is guaranteed to improve the performance of keyword queries.
   F. None of the above.

2. A C

3. (1 point) In vector-space model of retrieval, which of the following is true:

   A. Retrieval model is based on quantifying the similarity between query and document.
   B. The score computed by the retrieval model can be greater than 1, but is never less than 0.
   C. The model takes into account the effects of different document length.
   D. It makes no assumptions about the distribution of terms in documents.
   E. None of the above.

3. A C

4. (1 point) Which of the following measures can be used to evaluate the performance of a search system when we have a non-binary relevance judgements:

   A. Discounted Cumulative Gain
   B. Precision and recall
   C. Mean Average Precision
   D. F1-Measure
   E. None of the above

4. A

5. (1 point) During our discussion of Language Modeling for IR, we identified some weaknesses in the use of Multinomial as the underlying generative model. Which all of the following is a weakness that can be mitigated by the use of multiple-Poisson model instead:

   A. Occurrence of a term is independent of other terms
   B. Occurrence of a term is independent of previous occurrences of the same term
   C. Lack of ability to model bursty occurrences of some terms in a document
   D. No control on the size of the generated document
   E. None of the above

5. D

6. (1 point) Which of the following a valid reason for using interpolated values on a P-R curve:

   A. Drawing saw-tooth patterns for each query is hard as opposed to drawing x-axis parallel lines one gets after interpolation
   B. It makes it possible to compute average precision at a given recall value across queries
   C. It makes it possible to compute average recall at a given precision value across queries
   D. It is possible to compute average precision at an unknown recall value for a query
   E. Makes it more accurate due to rounding effects
   F. None of the above

6. B

7. (1 point) Which of the following is true:

   A. Typically webpages contain outgoing links to webpages whose URLs are closeby in lexicographic order
   B. Typically webpages contain outgoing links to webpages whose URLs have similar structure
   C. Typically webpages whose URLs share common prefixes have same outgoing links
   D. Typically webpage whose URLs share common prefixes have many common outgoing links
   E. None of the above

7. A D

8. So far we have seen a number of different retrieval models, which have conceptually distinct starting points, yet arrive at seemingly similar formulations. Despite these similarities, in practice each of them tends to perform differently. One way to reason about this behavior is to make a list of desirable properties of a good ranking function, and measure the extent to which these models satisfy those properties. For example, we may discover that one of the models satisfies a property only when documents are all of equal length, or when the queries consist of only rare terms etc.
Your instructor has suggested that the following are some of the desirable properties of a good ranking function:

> 1. All things being equal, ranking function should prefer documents containing more occurrences of a query term. i.e., if d is a document, and $q \in Q$ is a query term, and $t \notin Q$, then $RSV_{d\cup\{q\}} > RSV_{d\cup\{t\}}$.
>
> 2. if $q \in Q$ is a query term, $t \notin Q$, then
>
> $$\left(RSV_{d\cup\{q\}\cup\{t\}} - RSV_{d\cup\{t\}\cup\{t\}}\right) > \left(RSV_{d\cup\{q\}\cup\{q\}} - RSV_{d\cup\{q\}\cup\{t\}}\right).$$
>
> That is, a ranking function should resist gaming document relevance by stuffing query keywords to the document.
>
> 3. A ranking function should prefer the documents that *cover* more distinct query terms.
> In other words, if Q is the query, and $q_1, q_2 \in Q$ are two query terms. Assume $|d_1| = |d_2|$ and $idf(q_1) = idf(q_2)$. If $f_{q_1,d_1} = f_{q_1,d_2} + f_{q_2,d_2}$ and $f_{q_2,d_1} = 0, f_{q_1,d_2} \neq 0, f_{q_2,d_2} \neq 0$ then $RSV_{d_1} < RSV_{d_2}$.
>
> 4. If d is a document, and Q is the query, s.t. $q_1 \in Q$, $q_2 \in Q$, $q_1 \in d$, $q_2 \in d$, and $idf(q_1) > idf(q_2)$ and $f_{q_1,d} \leq f_{q_2,d}$, then $RSV_{d\cup\{q_1\}} > RSV_{d\cup\{q_2\}}$.
>
> 5. If d is a document, Q is a query, $d \cap Q \neq \emptyset$, and $d' = d \oplus d \oplus \ldots \oplus d$ is a document formed by concatenating document d with itself many times, then $RSV_{d'} \leq RSV_d$.

(in the above, $f_{w,d}$ is the frequency of term w in document d; and $idf(w)$ is the inverse document frequency of a term w)
Being a diligent IR student, you are not willing to just accept whatever your instructor says. You suspect that at least one of these properties as given above are not well formed. Also, you would like to conduct a diagnostic evaluation of different ranking functions you have learnt against these desirable properties. There are two tasks ahead of you:

(a) (5 points) Verify whether the above given properties are valid / invalid as one of the desirable properties of a ranking function – you can use any illustrative examples if needed. If any one of them is invalid, present its valid form (if possible).

(b) (6 points) Consider *vector-space models* and *Query likelihood LM with J-M smoothing.* Check which of the properties are satisfied by these ranking formulations and under which conditions. In all cases, clearly write down the representations for term-frequency and inverse-document frequency you are using, and the ranking formula used. If you are making any assumptions on the values taken by tf and idf values, state them and explain why they are realistic.

9. (2 points) When can the RSJ equation for RSV components for a term in the query become negative? Is that situation likely to occur in practice - explain.

10. (2 points) All the models we discussed so far were all based on the probability ranking principle (PRP) —i.e., given $P(R = 1|d, q)$ and $P(R = 0|d, q)$, rank documents in their estimated probability of relevance). While it has served well, there are settings where the PRP in the form as defined usually is not suitable to use. List at least two scenarios where retrieval has to move away from PRP, and provide arguments to support your claim.

11. Assume than an information retrieval engine returns a ranked list of 10 total documents for a given query. For a specific query assume that there are 5 relevant documents, but the resulting ranked list has relevant results only in positions 4, 6, 9.

    (a) (3 points) Compute the precision and recall values at each relevant document position and write them out as a table where each row contains the `<position, precision, recall>` (in that order).

12. (2 points) It is often observed that a search system has an undesirable property of exhibiting extreme diversity in its performance for different queries. In other words, it is not robust.

Consider two search systems, A and B, which are being evaluated against the same benchmark (consisting of a corpus, set of information need / queries, and qrels). We observe that System A has better MAP score than System B. Does this also mean System A is definitely more robust than System B? Explain clearly your reasoning steps. If needed, develop alternatives to MAP and / or identify other IR performance methods which are likely to give higher importance to robust systems.