

Indice

1	Warm up	2
2	Probability problems	3
2.1	Easy	3
3	Statistical problems	4
3.1	Easy	4
4	ML problems	5
4.1	Easy	5
5	Calculus / Algebra problems	7
5.1	Easy	7
6	Optimization problems	8
6.1	Easy	8
7	Algorithm problems	8
7.1	Easy	8
8	Miscellaneous problems	9
8.1	Easy	9

1 Warm up

- Prob You are presented with two empty jars and 100 marbles on a table. There are 50 white marbles and 50 black marbles. You are to put all 100 of the marbles into the two jars in any way you choose. I will then blindfold you. I will shake the jars up to ensure good mixing, and I will rearrange the placing of the jars on the table so that you do not know which one is which. You may then request either the "left-hand" or the "right-hand" jar. You get to choose exactly one jar, you are allowed to withdraw at most one marble from the jar, and you do not get a second chance if you are unhappy with your choice. How many of each color marble should you place in each jar to maximize the probability that your blindfolded random draw obtains a white marble?
- Prob Let's play a game. There are four sealed boxes. There is 100 pounds in one box and the others are empty. A player can pay X to open a box and take the contents as many times as they like. Assuming this is a fair game, what is the value of X ?
- Stat Explain collinearity and why this is a problem.
- ML How would you explain overfitting vs underfitting? And how should they detect which one is happening (as in, can you prescribe a test)?
- ML Why don't we just initialize all weights in a neural network to zero?
- Calc (Hard) What does derivative represent?
- Algo A group of people wants to determine their average salary on the condition that no individual would be able to find out anyone else's salary. Can they accomplish this, and, if so, how?

2 Probability problems

2.1 Easy

- HF. Let X, Y standard and independent gaussian random variables. Compute $\mathbb{E}[X+Y|X=2]$.
- HF. Let X a uniform random variable on $(0, \pi)$. Compute $\mathbb{E}(X|\sin X)$.
- HF. Let X_1, \dots, X_n uniform random variables on $(0, 1)$. Compute $\mathbb{E}(\max X_i)$.
- HF. Let X, Y standard gaussian random variables with correlation ρ . Compute $\mathbb{E}[X|Y]$.
- HF. Bubble sort is an algorithm that at each pass swaps the first and second index, if necessary, then the second and third, if necessary, and so on. It takes n passes in general. What is the probability that after one pass the array is sorted?
- HF. Let X, Y, Z three random variables with correlation ρ . Can you find a sharp inequality for ρ ?
20. One hundred people line up to board an airplane. Each has a boarding pass with assigned seat. However, the first person to board has lost his boarding pass and takes a random seat. After that, each person takes the assigned seat if it is unoccupied, and one of unoccupied seats at random otherwise. What is the probability that the last person to board gets to sit in his assigned seat?
10. For a 3 sets tennis game, would you bet on it finishing in 2 sets or 3 sets?
10. You have 17 coins and I have 16 coins, we flip all coins at the same time. If you have more heads then you win, if we have the same number of heads or if you have less then I win. What's your probability of winning?
- HS. You roll a die until a number other than a one appears. When such a number appears for the first time, I pay you the same number of dollars as there are dots on the upturned face of the die, and the game ends. What is the expected payoff to this game?
- HS. You are presented with two empty jars and 100 marbles on a table. There are 50 white marbles and 50 black marbles. You are to put all 100 of the marbles into the two jars in any way you choose. I will then blindfold you. I will shake the jars up to ensure good mixing, and I will rearrange the placing of the jars on the table so that you do not know which one is which. You may then request either the "left-hand" or the "right-hand" jar. You get to choose exactly one jar, you are allowed to withdraw at most one marble from the jar, and you do not get a second chance if you are unhappy with your choice. How many of each color marble should you place in each jar to maximize the probability that your blindfolded random draw obtains a white marble?
- MD. Flip 98 fair coins and 1 HH coin and 1 TT coin. Given that you see an H, what is the probability that it was the HH coin?

- MD. What is the probability of getting exactly 500 heads out of 1000 coin flips? Approximate it to within 5% of the true value.
- MD. Say you roll a die, and are given an amount in dollar equal to the number on the die. What would you pay to play this game if you played it a many times in a row? Now say that when you roll the die, you're allowed to either take the money that you'd get with the roll, or roll a second time; if you roll a second time, you're obligated to take the number of dollars that you get with the second roll. Now what is the worth of the game? Same thing as above, except you have an option to play the game a third time.
- MD. Let's say you're playing a two-player game where you take turns flipping a coin and whoever flips heads first wins. If the winner gets 1 dollar, how much would you pay to go first instead of second?
- QJ. Let's play a game. There are four sealed boxes. There is 100 pounds in one box and the others are empty. A player can pay X to open a box and take the contents as many times as they like. Assuming this is a fair game, what is the value of X ?
- QJ. We play a game: I pick a number n from 1 to 100. If you guess correctly, I pay you n and zero otherwise. How much would you pay to play this game?
- QJ. Suppose you are throwing a dart at a circular board. What is your expected distance from the center? Suppose you win a dollar if you hit 10 times in a row inside a radius of $R/2$, where R is the radius of the board. You have to pay 0.10 dollar for every try. If you try 100 times, how much money would you have lost/made on average? Does your answer change if you are a professional and your probability of hitting inside $R/2$ is double of hitting outside $R/2$?
- AP. How many people do we need in a class to make the probability that two people have the same birthday more than $1/2$? (For simplicity, assume 365 days a year.)
- AP. In a primitive society, every couple prefers to have a baby girl. There is a 50% chance that each child they have is a girl, and the genders of their children are mutually independent. If each couple insists on having more children until they get a girl and once they have a girl they will stop having more children, what will eventually happen to the fraction of girls in this society?
- GR. Suppose that X and Y are mean zero, unit variance random variables. If least squares regression (without intercept) of Y against X gives a slope of β , what is the slope of the regression of X against Y ?

3 Statistical problems

3.1 Easy

- HF. Derive the OLS equations for a dataset (X, Y) with $n = 1$ and for a generic n . What are the issues with this equation?

- HF. Assuming I have n data points and k features, with $k > n$ and I want to create a OLS model? How can I solve the problems from $k > n$?
- MD. Explain collinearity and why this is a problem.
- QT. What is the Law of Large Numbers? The Central Limit Theorem? Give an example(s) of a distribution that satisfies one, but not both.
- QJ. I have a bag containing nine ordinary coins and one double-headed one. I remove a coin and flip it three times. It comes up heads each time. What is the probability that it is the double-header?
- QJ. I take an ordinary-looking coin out of my pocket and flip it three times. Each time it is a head. What do you think is the probability that the next flip is also a head? What if I had flipped the coin 100 times and each flip was a head?
- DP. Describe MLE and where we need this.
- HU. It's a common practice to assume an unknown variable to be of the normal distribution. Why is that?
- HU. A/B testing is a method of comparing two versions of a solution against each other to determine which one performs better. What are some of the pros and cons of A/B testing?
- HU. How do you assess the statistical significance of a pattern whether it is a meaningful pattern or just by chance?

4 ML problems

4.1 Easy

- HF. What's your favourite machine learning model? Where is it applied? Major issues and how did the community solved them?
- HF. Explain the use of gradient descent. Give a formal definition of what gradient descent is.
- LE. Implement a forward and backward pass for a two-layer NN.
- LE. How would you explain overfitting vs underfitting? And how should they detect which one is happening (as in, can you prescribe a test)?
- LE. Can you please plot "learning curves" for a typical ML system? And can you specialize these plots for the overfitting vs underfitting cases?
- LE. Can you explain k -fold cross-validation? And why/when do we use it?
- DP. What is the curse of dimensionality?

- DP. Name three possible sources of uncertainty in ML?
- DP. Why do we need parameter sharing?
- HU. Explain supervised, unsupervised, weakly supervised, semi-supervised, and active learning.
- HU. If we have a wide NN and a deep NN with the same number of parameters, which one is more expressive and why?
- HU. What are saddle points and local minima? Which are thought to cause more problems for training large NNs?
- HU. Why is hyperparameter tuning important? Do you know any algorithm to tune hyperparameters?
- HU. What makes a classification problem different from a regression problem? Can you turn one problem to another?
- HU. How to determine whether two sets of samples (e.g. train and test splits) come from the same distribution?
- HU. How would class imbalance affect your model? Why is it hard for ML models to perform well on data with class imbalance? Imagine you want to build a model to detect skin lesions from images. In your training dataset, only 1% of your images shows signs of lesions. After training, your model seems to make a lot more false negatives than false positives. What are some of the techniques you'd use to improve your model?
- HU. Why does normalization help prevent feature leakage? How do you detect feature leakage?
- HU. When we say an algorithm converges, what does convergence mean? How do we know when a model has converged?
- HU. What's the bias-variance trade-off? How's this tradeoff related to overfitting and underfitting? How do you know that your model is high variance, low bias? What would you do in this case? What about the opposite?
- HU. What are the basic assumptions to be made for linear regression?
- HU. What happens if we don't apply feature scaling to logistic regression?
- HU. Why do we use feature selection?
- HU. Your client wants you to train a language model on their dataset but their dataset is very small with only about 10,000 tokens. Would you use an n-gram or a neural language model?
- HU. What is the role of zero padding?
- HU. What happens when you use max-pooling instead of average pooling?

HU. You want to build a classifier to predict sentiment in tweets but you have very little labeled data (say 1000). What do you do?

HU. When building a neural network, should you overfit or underfit it first?

HU. How do we know that gradients are exploding? How do we prevent it? Why are RNNs especially susceptible to vanishing and exploding gradients?

HU. Some models use weight decay: after each gradient update, the weights are multiplied by a factor slightly less than 1. What is this useful for?

HU. Why don't we just initialize all weights in a neural network to zero?

HU. What's a dead neuron? How to detect and prevent them?

5 Calculus / Algebra problems

5.1 Easy

HF. Compute $\int_0^\infty e^{-x} dx$.

HF. Is e^π greater than π^e ?

HF. Find a lower bound and an upper bound for $\sqrt{107}$.

HF. Invert a 2×2 matrix.

HF. Find eigenvalues and eigenvectors of a 2×2 matrix.

QT. Suppose M is an $n \times n$ correlation matrix, with correlation ρ between any pair of random variables. What is the smallest possible value of ρ ?

HU. What's the geometric interpretation of the dot product of two vectors? Given a vector u , find vector v of unit length such that the dot product of u and v is maximised.

HU. What does it mean for two vectors to be linearly independent?

HU. Why do we say that matrices are linear transformations?

HU. What's the inverse of a matrix? Do all matrices have an inverse? Is the inverse of a matrix always unique?

HU. What does the determinant of a matrix represent? What happens to the determinant of a matrix if we multiply one of its rows by a scalar?

HU. What does derivative represent?

6 Optimization problems

6.1 Easy

HF. Given a parallelepiped of size x, y, z and a given volume xyz , you want to minimize the lateral area and the base area.

HF. Let x_1, \dots, x_n real numbers. Find x such that $\sum_{i=1}^n |x - x_i|$ is minimized.

HF. Let a, b real numbers such that $a + 2b \leq 50$. Find the maximum of ab .

7 Algorithm problems

7.1 Easy

HF. Find an efficient algorithm to compute the square root of y . What's the convergence time? How can I choose the starting x_0 value.

LE. Given a list L and an integer k , select k elements from L without replacement.

LE. What is a good storage strategy for sparse vectors such that we get fast dot product?

LE. Given a collection of n two dimensional points and a number k , return the k closest points to $(0, 0)$ by Euclidean distance.

10. A group of people wants to determine their average salary on the condition that no individual would be able to find out anyone else's salary. Can they accomplish this, and, if so, how?

10. How many digits are in 99 to the 99th power?

QT. Explain how a hash table works.

QT. You are given an unsorted list of 999,000 unique integers, each from 1 and 1,000,000. Find the missing 1000 numbers. What is the computational complexity of your solution?

HU. You have three matrices: A, B, C with dimensions $100 \times 5, 5 \times 200, 200 \times 20$ respectively and you need to calculate the product ABC . In what order would you perform your multiplication and why?

HU. What are some of the causes for numerical instability in deep learning?

8 Miscellaneous problems

8.1 Easy

- HF. How can you assign n chairs to k people such that for every two people there is at least an empty chair?
- HF. We call *beautiful* a binary sequence where two consecutive 1 don't appear. Compute the number of beautiful sequences with length k .
- AP. Suppose we have 98 distinct integers from 1 to 100. What is a good way to find out the two missing integers (within $[1, 100]$)?