0/12 Questions Answered

2021Final

Choose Files No file chosen

Q1 PCA
5 Points

Q1.1 Singular values
1 Point

Let X be a $n \times n$ symmetric matrix whose eigenvalues are $\lambda_1, \ldots, \lambda_n$. What are the singular values of the matrix $Y = I - X$, where I denotes the $n \times n$ identity matrix.

Enter your answer here

Save Answer

Q1.2 PCA
1 Point

Describe how you might use PCA to do dimension reduction as a pre-processing step when dealing with data of the form $x_1, \ldots, x_n \in R^d$.

Enter your answer here

Save Answer

Q1.3 Powering
3 Points

Let A, B be two sparse $n \times n$ matrices with A, B having p-fraction of non-zeros and let U be a $n \times k$ matrix. Let $C = A^4 - 2A^3B + B^4 + AUU^TB^T$.

Give an algorithm based on power iteration to compute the top singular vector of C.

For full credit, your algorithm should not have to compute the matrix C.

Enter your answer here

Choose Files No file chosen

Please select file(s) Select file(s)

Save Answer

Q2 Optimization

5 Points

Q2.1 GD vs SGD

1 Point

What is the computational advantage of Stochastic gradient descent over gradient descent as applied to empirical risk minimization (ERM)?

Enter your answer here

Save Answer

Q2.2 Convexity

1 Point

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Is $g(x) = f(-x)$ convex? (Justify your answer.)

Enter your answer here

Save Answer

Q2.3 Quant at a hedge fund

3 Points

Suppose you are a QA at a hedge fund with M amount of funds and have to decide on what fraction of the fund to invest in each of d asset classes (e.g., different kinds of stock, cryptocurrency, bonds). Each of these asset classes has a predicted profit margin that you know (or willing to work with): One dollar invested in class i returns a profit of m_i dollars.

If this were the only information you had, then you'd invest completely in the

Choose Files No file chosen

asset with the best profit. However, this is much riskier than using multiple assets. To model this, suppose you also want to account for how skewed your investments are. Define the net risk of investing x dollars in asset class i as $r_i x^2$, where $r_i > 0$ is a parameter that you know (or have an estimate of). The total risk is the sum of risk across assets.

Write down an optimization problem to minimize $risk - profit$ (the totals). What algorithm would you use to optimize this function? Write down the algorithm.

Enter your answer here

Please select file(s) Select file(s)

Save Answer

Q3 The Good Experts

5 Points

Consider learning with experts scenario as in class where each expert incurs losses that are either 0 or 1. We saw in class that the MWM can get a regret bound of $O(\sqrt{T \log d})$ where d is the number of experts.

Now, suppose that we are in a scenario, where there is not just one best expert, but a noticeable fraction of them have good performance. For instance, if all experts are great, then we should expect to do better. Can MWM exploit this to get a better performance?

Suppose that for some $\gamma \in (0, 1)$, at least γ fraction of the experts incur total losses of at most L. That is, if $L(t, i)$ denotes the loss of expert i in round t (as in class), then $|\{i : \sum_{t=1}^{T} L(t, i) \leq L\}| \geq \gamma d$. Show that MWM run with step-size ϵ (i.e., changing weights by multiples of $(1 - \epsilon)$ as in class) incurs a total loss of at most $(1 + \epsilon)L + \ln(1/\gamma)/\epsilon$. Conclude, that we can get a regret of $O(\sqrt{L \ln(1/\gamma)})$.

Enter your answer here

Please select file(s) Select file(s)

Choose Files No file chosen
Save Answer

Q4 Graphical Models
5 Points

Q4.1 Independence vs Dependence
2 Points

Give an example of random variables (X_1, X_2, X_3, X_4) where any three are independent of each other but together they are dependent.

Enter your answer here

Please select file(s) Select file(s)

Save Answer

Q4.2 GLM via Sparsitron
3 Points

In class we saw the sparsitron algorithm to solve GLM where our goal was to minimize loss of the form $E[(\sigma(p \cdot X) - \sigma(p_* \cdot X))^2]$, from samples of the form (X, Y), where $E[Y|X = x] = \sigma(p_* \cdot X)$. When we discussed the algorithm inc lass, we only wrote it down for the case where we assumed that p_* was non-negative and norm 1. That is, $\sum_i (p_*)_i = 1$. We also described ways to get around this assumptions.

Put these together, and write down the general algorithm for GLM minimization: We see samples of the form (X, Y) where $E[Y|X = x] = \sigma(p_* \cdot X)$ just as in class. But we only know that $\sum_i |(p_*)_i| = \lambda$ for a parameter λ that we know. It might be easiest to write down the pseudocode, but you can also explain what can be done.

No need to analyze the algorithm.

Enter your answer here

Please select file(s) Select file(s)

Choose Files No file chosen

Save Answer

Q5 Streaming algorithms

5 Points

Q5.1 Frequency estimation

2 Points

Consider a datastream $x = (x_1, \ldots, x_n)$ with $x_i \in \{1, \ldots, m\}$ and let $f = (f_1, \ldots, f_m)$ be the associated frequency vector. The goal of this problem is to compute the squared-norm of the frequency vector $||f||_2^2 = \sum_{a=1}^{m} f_a^2$. Consider the following algorithm:

1. Initialize $C = 0$. Choose a random string $z \in \{1, -1\}^m$ (each coordinate z_a is independent and uniform in $\{1, -1\}$).
2. For each $i = 1, \ldots, n$:
   a. Set $C = C + z_{x_i}$.
   b. Output $g(C)$ for some univariate function $g : R \to R$.

What should g be so that $E[g(C)] = ||f||_2^2$? (The expectation is over the random string z.)

(Note that $g : R \to R$ is a fixed function that only gets C as input and nothing else.)

Enter your answer here

Please select file(s) Select file(s)

Save Answer

Q5.2 Estimator correctness

1 Point

Show that the function g you defined in the previous question satisfies the required property: That is, prove that $E[g(C)] = \|f\|_2^2$.

Enter your answer here

Please select file(s) Select file(s)

Save Answer

Q5.3 Multiple-Counts

2 Points

In class we designed frequency estimation algorithms where the stream was of the form $(x_1, \ldots, x_n)$ with $x_i \in \{1, \ldots, m\}$. Let us consider a more general scenario where the stream is of the form $((x_1, c_1), (x_2, c_2), \ldots, (x_n, c_n))$ where for $i = 1, \ldots, n$, x_i is an element of $\{1, \ldots, m\}$ and $c_i > 0$ is an integer to be interpreted as the number of copies of x_i that are being added to the stream. In particular, the frequency vector of the stream would be $f = (f_1, \ldots, f_m)$ where $f_a = \sum_{i:x_i=a} c_i$. Extend Count-Min sketch to give a frequency estimation algorithm for this more general setup.

(For full credit, it suffices to specify the algorithm to process the stream and the query algorithm; you don't have to analyze them. Also, for full credit, your update times have to be of the same order as the algorithm we did in class.)

Enter your answer here

Please select file(s) Select file(s)

Save Answer

Save All Answers

Submit & View Submission