Problem 1:

Communications systems are a complex thing to generalize mathematically. In this paper , “A Mathematical Theory of Communication” a communication system is comprised of five components. The first is an information source, which is the origin of the message. The second is the transmitter, or producer of the signal to be carried on a channel. The third is the channel (think wires, radio frequencies, etc.). Next is the receiver which performs the inverse operation as the transmitter. Last is the destination where the message was intended to arrive.

The required capacity of a channel in a communication system can be simplified by using statistics. The paper explains that certain letters and combinations of letters are more common than others. This is used to decrease the amount of data that needs to flow across a system. This is the same concept used in what is called a stochastic process, which is defined as “A physical system, or a mathematical model of a system which produces such a sequence of symbols governed by a set of probabilities”. A stochastic process can be a written language of some sort, or anything that generates a series of characters.

The paper gets particularly interesting when it starts talking about the generation of strings based purely on statistics. It talks about zero-order approximation, which is simply randomly selecting characters. Naturally, this results in gibberish. Once we start introducing statistics, using the probability of each character (first-order approximation), we still get gibberish, but it starts to kind of look like the words might be English. If we move to second-order approximation, which is using probabilities of a letter using the letter that preceded it, it becomes more convincing. By as early as third-layer approximation we start getting real words.

Another notable mention in this paper is the concept of entropy. Entropy is the probability that a randomly selected word is a given word. This is quite simple when you are looking at a single word, but gets to be quite computationally expensive when using n-layer approximation.

Problem 2:

NOTE: I used documents from the page that had the words “Download PDF” or “Supplementary PDF”. This resulted in 1087 PDFs.

1. Top ten most common words:
can, learning, model, function, set, data, algorithm, using, number, also
This was done removing stopwords, any words that were not part of the English language, numbers, empty strings, single-character words, and any punctuation.

2. Using sum(-pi * log(pi)), The entropy I get is 7.9591.

3. Supplement multiplication intersection weakly componentwise rooting Rob particular Assumption directly gradient consider can make number model However optimization smoothing number Since Bayesian performance base congruent also

Knowledge increasing verify unconstrained avoid hence algorithm twice random machine accuracy arg point Computer thank addition layer end Repeated guarantees neural Moreover dense can magnitude im information end Lemma eq issue Deep choose add converge throughout network TRACE rate will orders Research forcing cf asked onto expected Since process label develop Gaussian exists tropical can real test policies make may live include blue gradients See data doubly Shamir avoid loss deep Figure method loss Output reward many diversity collection see spatial quantitative data grants converged learning decision adaptive guarantee algorithms Theorem terminates number fared recommended solutions sharp algorithm diffraction joint calculate Property Let BN hold utilizing models can standard equalized places nuclear performances increase treatment observes helpful similar min ran accelerated gradient evaluation density connected books Proceedings sequences post covariates results increasing Algorithm expected Unsupervised Size Gaussian prob robust variance state saddle neural corresponding thus Advances Nonparametric objective scalar local work Let sub augmented term learning performance grange inference compute prob Figure common loo domain optimization given entropy intelligence parameter focusing Let kernel geometric

4. For my paragraph creation, every previous word was picked depending on the word the preceded it where possible. If I had more time, I would improve it by further cleaning up my word selection when reading the PDFs. I would also use punctuation to unlink unrelated words at the end of a sentence. Lastly, I would clean up all of the variable words that I introduced from the n-gram generation. Below is my paragraph.

Sparse inverse estimation problem of requires solving the orange or Assumption there selecting an important quantity for return of sentences taken together segments from L steps to Since we application of constraint that task increase as prohibitively expensive a amounts to pixel could Cheng et al show that IEEE transactions on performance is require it investigator is Coding proof to al show that opposed to additive theorem examine the without any invertible matrix while th international conference on translate into matrix while loop is eT with Combinatorial bandits which Note that different coordinates the generation scheme that better update schemes for utility typically be black box constraints siam solving the se samples from tor based on Implementation and tree of O contributions as ter update schemes for Astate the performance is pi li ui and radius example in compute a around nonzero entries set the student and avoids the descent exploration the However the elements are actual inference for solved using the al show that cross validation to International conference on feedback possibly a controlled by metric for S action a posed method that Next we property for regularizing matrix while recent results to privacy as t the rules finally fuels figure xR rd rs coordinate is forest etc extrapolated are P v dataset into single hidden layer of d dimensions is propose and color ordinal or training data suppose the practice see figure Learning thomas Conclusion and image processing systems pp a Raghu et al show that reconstruct a entropy and L steps to attentive in lower dimensionality dx and prior work et al show that yi where start state variables and London conn comp rec systems pp a gradient algorithms we ensures that used updates for control of