Labs
Machine Learning Course
Fall 2025

Westlake University
Department of Artificial Intelligence, SOE
Tao Lin
https://github.com/LINs-lab/course_machine_learning

Problem Set Lab 01 (graded), Sept. 04, 2025 (Mathematical Foundation of Machine Learning)

Goals. The goals of this lab are to:

- Familiarize yourself with the mathematical foundations of the machine learning course.

Submission instructions:

- Please submit a PDF file to canvas.
- Deadline: 23.59 on Mar. 02, 2025

Review of Linear Algebra

Problem 1 (Idempotent Matrices and Rank Inequality):

Given A and B are idempotent $n \times n$ matrices (i.e., $A^2 = A$ and $B^2 = B$), and they commute with each other ($AB = BA$):

1. Prove that $A + B - AB$ is also an idempotent matrix.
2. Further prove that

$$\mathrm{rank}(A + B - AB) \leq \mathrm{rank}(A) + \mathrm{rank}(B)\,.$$

Solution 1 (Idempotent Matrices and Rank Inequality):

Let A and B be idempotent matrices ($A^2 = A, B^2 = B$) such that they commute ($AB = BA$).

Part 1: Prove that $A + B - AB$ is idempotent.

We aim to show that $(A + B - AB)^2 = A + B - AB$. Let's expand the square:

$$\begin{aligned}
(A + B - AB)^2 &= (A + B - AB)(A + B - AB) \\
&= A(A + B - AB) + B(A + B - AB) - AB(A + B - AB) \\
&= A^2 + AB - A^2B + BA + B^2 - B^2A - ABA - AB^2 + ABAB
\end{aligned}$$

Using the given properties $A^2 = A$, $B^2 = B$, and $AB = BA$, we can simplify the expression:

$$\begin{aligned}
&= A + AB - AB + AB + B - BA - A(BA) - AB + A(BA)B \\
&= A + AB + B - AB - A(AB) - AB + A(AB)B \\
&= A + B - A^2B - AB + A^2B^2 \\
&= A + B - AB - AB + AB \\
&= A + B - AB
\end{aligned}$$

Thus, the matrix $A + B - AB$ is idempotent. This completes the first part of the proof.

Part 2: Prove the rank inequality.

We aim to show that $\mathrm{rank}(A+B-AB) \leq \mathrm{rank}(A) + \mathrm{rank}(B)$. A key property of any idempotent matrix M is that its rank is equal to its trace, i.e., $\mathrm{rank}(M) = \mathrm{tr}(M)$. This holds because the eigenvalues of an idempotent matrix can only be 0 or 1, and the rank is the number of non-zero eigenvalues (counting multiplicity), which equals their sum (the trace).

Since A, B, and (from Part 1) $A + B - AB$ are all idempotent, we have:

$$\begin{aligned}
\mathrm{rank}(A) &= \mathrm{tr}(A) \\
\mathrm{rank}(B) &= \mathrm{tr}(B) \\
\mathrm{rank}(A + B - AB) &= \mathrm{tr}(A + B - AB)
\end{aligned}$$

By the linearity of the trace operator:

$$\mathrm{tr}(A + B - AB) = \mathrm{tr}(A) + \mathrm{tr}(B) - \mathrm{tr}(AB)$$

Since A and B are commuting idempotent matrices, they are simultaneously diagonalizable. Their eigenvalues are in $\{0, 1\}$. Therefore, the eigenvalues of their product AB are also in $\{0, 1\}$, which implies $\mathrm{tr}(AB) \geq 0$.

It follows that:

$$\begin{aligned}
\mathrm{rank}(A + B - AB) &= \mathrm{tr}(A + B - AB) \\
&= \mathrm{tr}(A) + \mathrm{tr}(B) - \mathrm{tr}(AB) \\
&\leq \mathrm{tr}(A) + \mathrm{tr}(B) \\
&= \mathrm{rank}(A) + \mathrm{rank}(B)
\end{aligned}$$

This completes the proof.

Problem 2 (Diagonalizability and Eigenvalues of a Linear Transformation):

Let V be a four-dimensional vector space, and let $T : V \to V$ be a linear transformation satisfying $T^3 - 2T^2 + T - 2I = 0$, where I is the identity transformation on V. Prove that T is diagonalizable and determine all its eigenvalues.

Solution 2 (Diagonalizability and Eigenvalues of a Linear Transformation):

Let the given **polynomial** be $p(x) = x^3 - 2x^2 + x - 2$. The linear transformation T satisfies $p(T) = 0$.

Part 1: Prove that T is diagonalizable.

A linear transformation is diagonalizable if and only if its minimal polynomial, $m_T(x)$, splits into distinct linear factors. By definition, the minimal polynomial $m_T(x)$ must divide any polynomial $p(x)$ for which $p(T) = 0$. Let us find the roots of $p(x)$:

$$\begin{aligned} p(x) = x^3 - 2x^2 + x - 2 &= x^2(x-2) + 1(x-2) \\ &= (x-2)(x^2+1) \\ &= (x-2)(x-i)(x+i) \end{aligned}$$

The roots of $p(x)$ are $2, i, -i$. Since these roots are distinct, any polynomial that divides $p(x)$, including the minimal polynomial $m_T(x)$, must also have distinct roots. Because $m_T(x)$ has distinct linear factors (over $\mathbb{C}$), the transformation T is diagonalizable. This completes the proof of diagonalizability.

Part 2: Determine all its eigenvalues.

The set of eigenvalues of a linear transformation T is precisely the set of roots of its minimal polynomial $m_T(x)$. As established, $m_T(x)$ must divide $p(x) = (x-2)(x-i)(x+i)$. Therefore, the roots of $m_T(x)$ must be a subset of the roots of $p(x)$. The problem implies a non-trivial transformation on a four-dimensional space, so the set of eigenvalues cannot be empty.

Thus, the set of eigenvalues of T, denoted $\sigma(T)$, must be a non-empty subset of $\{2, i, -i\}$. The specific characteristic polynomial depends on the multiplicities of these eigenvalues, which must sum to the dimension of the vector space, 4.

Problem 3 (Existence of Real Matrix Roots for Positive Eigenvalue Matrices):

If a real matrix A has all eigenvalues as positive real numbers, then for any positive integer m, there exists a real matrix B such that $B^m = A$.

Solution 3 (Existence of Real Matrix Roots for Positive Eigenvalue Matrices):

Since all eigenvalues of the real matrix A are positive, the matrix logarithm $\log(A)$ is well-defined and is a real matrix. This can be constructed via the power series for $\log(I + X)$ or by applying the logarithm function to the Jordan blocks of A. Since eigenvalues are positive, the function is defined on the spectrum of A.

Let us define the matrix B using the matrix exponential:

$$B = \exp\left(\frac{1}{m}\log(A)\right)$$

Since $\log(A)$ is a real matrix, and m is a positive integer, $\frac{1}{m}\log(A)$ is also a real matrix. The matrix exponential of any real matrix is real, therefore B is a real matrix.

We now verify that $B^m = A$. Using the **(Matrix Exponential)** property that $(\exp(X))^m = \exp(mX)$ for any matrix X:

$$\begin{aligned}
B^m &= \left[\exp\left(\frac{1}{m}\log(A)\right)\right]^m \\
&= \exp\left(m \cdot \frac{1}{m}\log(A)\right) \\
&= \exp(\log(A)) \\
&= A
\end{aligned}$$

Thus, we have constructed a real matrix B such that $B^m = A$. This completes the proof.

Problem 4 (Eigenvalue Equivalence under Commutator-like Condition):
Let A and B be $n \times n$ square matrices satisfying

$$AB - BA = A - B\,.$$

Then, A and B have the same eigenvalues.

Solution 4 (Eigenvalue Equivalence under Commutator-like Condition):

Let $\sigma(A)$ and $\sigma(B)$ denote the set of eigenvalues (the spectrum) of A and B, respectively.

First, we establish a key implication. Let λ be an eigenvalue of A. This means there exists a non-zero eigenvector v such that $Av = \lambda v$, or equivalently, $(A - \lambda I)v = 0$.

The given relation $AB - BA = A - B$ can be rewritten by subtracting $\lambda B - \lambda B$ and $\lambda A - \lambda A$, and rearranging:

$$\begin{aligned} AB - BA - (A - B) &= 0 \\ (AB - \lambda B) - (BA - \lambda B) - (A - \lambda I) + (B - \lambda I) &= 0 \\ (A - \lambda I)B - B(A - \lambda I) &= (A - \lambda I) - (B - \lambda I) \end{aligned}$$

Applying this operator equation to the eigenvector v:

$$\begin{aligned} ((A - \lambda I)B - B(A - \lambda I))\,v &= ((A - \lambda I) - (B - \lambda I))\,v \\ (A - \lambda I)Bv - B\,\underbrace{(A - \lambda I)v}_{=0} &= \underbrace{(A - \lambda I)v}_{=0} - (B - \lambda I)v \\ (A - \lambda I)Bv &= -(B - \lambda I)v \end{aligned}$$

We now consider two cases.

1. **Case 1:** $(B - \lambda I)v = 0$.
   Since v is a non-zero vector, this equation directly implies that λ is an eigenvalue of B.

2. **Case 2:** $(B - \lambda I)v \neq 0$.
   Let's define a new non-zero vector $u = (B - \lambda I)v$. From the definition of u, we have $Bv = u + \lambda v$. Substituting this into our derived relation $(A - \lambda I)Bv = -u$ gives:

$$\begin{aligned} (A - \lambda I)(u + \lambda v) &= -u \\ (A - \lambda I)u + \lambda(A - \lambda I)v &= -u \\ (A - \lambda I)u + \lambda(0) &= -u \\ Au - \lambda u &= -u \\ Au &= (\lambda - 1)u \end{aligned}$$

   Since $u \neq 0$, this shows that $\lambda - 1$ is an eigenvalue of A.

Combining these two cases, we have established the first property:

If $\lambda \in \sigma(A)$, then either $\lambda \in \sigma(B)$ or $\lambda - 1 \in \sigma(A)$.

The original equation is equivalent to $BA - AB = B - A$. The relationship is symmetric upon swapping A and B and negating the equation. Therefore, an identical argument yields the symmetric property:

If $\mu \in \sigma(B)$, then either $\mu \in \sigma(A)$ or $\mu - 1 \in \sigma(B)$.

Assume, for the sake of contradiction, that $\sigma(A) \neq \sigma(B)$. This means that at least one of the sets $\sigma(A) \setminus \sigma(B)$ or $\sigma(B) \setminus \sigma(A)$ is non-empty.

Let's suppose $\sigma(A) \setminus \sigma(B)$ is non-empty. Choose any eigenvalue $\lambda_0 \in \sigma(A) \setminus \sigma(B)$. We know $\lambda_0 \in \sigma(A)$ and $\lambda_0 \notin \sigma(B)$. By the 1st property above, it must be that $\lambda_0 - 1 \in \sigma(A)$. Now consider $\lambda_1 = \lambda_0 - 1$. If λ_1 is also not in $\sigma(B)$, then applying the same property again shows that $\lambda_1 - 1 = \lambda_0 - 2$ must also be in $\sigma(A)$.

This process can be repeated. As long as the new eigenvalue is not in $\sigma(B)$, we generate another, smaller eigenvalue of A. This would create an infinite sequence of distinct eigenvalues for A: $\lambda_0, \lambda_0 - 1, \lambda_0 - 2, \ldots$.

This is a contradiction, because an $n \times n$ matrix has a characteristic polynomial of degree n and can therefore have at most n distinct eigenvalues.

Our assumption that $\sigma(A) \setminus \sigma(B)$ is non-empty must be false. Thus,

$$\sigma(A) \subseteq \sigma(B)$$

.

By a perfectly symmetric argument using the property from Step 2, the set $\sigma(B) \setminus \sigma(A)$ must also be empty. Thus,

$$\sigma(B) \subseteq \sigma(A)$$

.

Since $\sigma(A) \subseteq \sigma(B)$ and $\sigma(B) \subseteq \sigma(A)$, we conclude that $\sigma(A) = \sigma(B)$.

Problem 5 (Matrix Determinant and Commutator):

Let A and B be two $n \times n$ matrices satisfying the equation

$$AB - BA = A.$$

Prove that $\det(A) = 0$.

Solution 5 (Matrix Determinant and Commutator):

Assume A is invertible, i.e., $\det(A) \neq 0$. Then, A^{-1} exists. Multiply both sides of the equation $AB - BA = A$ on the left by A^{-1}

$$A^{-1}AB - A^{-1}BA = A^{-1}A\,.$$

Simplifying, we get

$$B - A^{-1}BA = I\,,$$

where I is the identity matrix. Take the trace of both sides:

$$\mathsf{tr}(B - A^{-1}BA) = \mathsf{tr}(I)\,,$$

Using the cyclic property of trace ($\mathsf{tr}(A^{-1}BA) = \mathsf{tr}(BAA^{-1}) = \mathsf{tr}(B)$), the LHS simplifies to

$$\mathsf{tr}(B) - \mathsf{tr}(B) = 0\,.$$

The RHS is

$$\mathsf{tr}(I) = n\,.$$

This leads to the contradiction

$$0 = n\,.$$

Since n is a positive integer, the assumption that A is invertible is false. Therefore, A is singular, which means:

$$\det(A) = 0\,.$$

Review of Probability Theory

Problem 6 (Moment Bound for a Standard Normal Random Variable):

For a standard normal random variable X, there exists a constant C such that for all $p > 1$,

$$(\mathbb{E}[|X|^p])^{1/p} \leq C\sqrt{p}\,.$$

Solution 6 (Moment Bound for a Standard Normal Random Variable):

For a standard normal random variable X, the p-th moment is given by:

$$\mathbb{E}[|X|^p] = 2^{p/2}\frac{\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}\,,$$

where Γ denotes the Gamma function. Stirling's approximation provides an asymptotic expression for the Gamma function for large arguments:

$$\Gamma(z) \sim \sqrt{2\pi}\, z^{z-1/2}e^{-z}, \quad \text{as } z \to \infty\,.$$

Let us set $z = \frac{p+1}{2}$. Then, for large p ,

$$\Gamma\left(\frac{p+1}{2}\right) \sim \sqrt{2\pi}\left(\frac{p+1}{2}\right)^{\frac{p}{2}} e^{-\frac{p+1}{2}}\,.$$

For simplicity, and since p is large, we can approximate $p + 1 \approx p$, yielding:

$$\Gamma\left(\frac{p+1}{2}\right) \sim \sqrt{2\pi}\left(\frac{p}{2}\right)^{\frac{p}{2}} e^{-\frac{p}{2}}\,.$$

Substituting the Stirling approximation into the expression for $\mathbb{E}[|X|^p]$:

$$\mathbb{E}[|X|^p] \sim 2^{p/2}\cdot\frac{\sqrt{2\pi}\left(\frac{p}{2}\right)^{\frac{p}{2}} e^{-\frac{p}{2}}}{\sqrt{\pi}} = 2^{p/2}\cdot\sqrt{2}\left(\frac{p}{2}\right)^{\frac{p}{2}} e^{-\frac{p}{2}}\,.$$

Simplifying further:

$$\mathbb{E}[|X|^p] \sim 2^{p/2}\cdot\sqrt{2}\cdot\left(\frac{p}{2}\right)^{\frac{p}{2}} e^{-\frac{p}{2}} = \sqrt{2}\left(\frac{p}{e}\right)^{\frac{p}{2}}\,.$$

Taking the p-th root of both sides:

$$(\mathbb{E}[|X|^p])^{1/p} \sim \left(\sqrt{2}\left(\frac{p}{e}\right)^{\frac{p}{2}}\right)^{1/p} = 2^{1/(2p)}\left(\frac{p}{e}\right)^{1/2}\,.$$

As $p \to \infty$, $2^{1/(2p)} \to 1$, so for sufficiently large p,

$$(\mathbb{E}[|X|^p])^{1/p} \lesssim \sqrt{\frac{p}{e}}\,.$$

Thus, there exists a constant $C \geq \frac{1}{\sqrt{e}}$ such that

$$(\mathbb{E}[|X|^p])^{1/p} \leq C\sqrt{p}\,.$$

The above asymptotic analysis holds for large p. For smaller values of p, the moments $\mathbb{E}[|X|^p]^{1/p}$ can be explicitly computed or bounded, and they are finite.

Problem 7 (Bounded Random Variable and Exponential Expectation):

Let X be a bounded random variable with $\mathbb{E}[X] = 0$ and $|X|_\infty \leq a$ for some $a > 0$. Prove that

$$\mathbb{E}\left[e^X\right] \leq \cosh(a)\,.$$

Solution 7 (Bounded Random Variable and Exponential Expectation):

Let X be a random variable such that $-a \leq X \leq a$ and $\mathbb{E}[X] = 0$. We want to find the maximum value of $\mathbb{E}[e^X]$.

The function $\phi(x) = e^x$ is convex. For any value $x \in [-a, a]$, it lies on the line segment connecting the points $(-a, \phi(-a))$ and $(a, \phi(a))$. Any such x can be written as a convex combination of the endpoints $-a$ and a:

$$x = \lambda a + (1 - \lambda)(-a) \quad \text{for some } \lambda \in [0, 1]\,.$$

Solving for λ, we find $x = 2\lambda a - a$, which gives $\lambda = \frac{x+a}{2a}$. Since $x \in [-a, a]$, we can verify that $\lambda \in [0, 1]$.

By the definition of convexity, for any $x \in [-a, a]$, we have:

$$\phi(x) \leq \lambda\phi(a) + (1 - \lambda)\phi(-a)$$

Substituting $\phi(x) = e^x$ and the expression for λ:

$$e^x \leq \left(\frac{x+a}{2a}\right)e^a + \left(1 - \frac{x+a}{2a}\right)e^{-a} = \left(\frac{x+a}{2a}\right)e^a + \left(\frac{a-x}{2a}\right)e^{-a}\,.$$

This inequality holds for any possible value of the random variable X. Therefore, we can take the expectation of both sides:

$$\mathbb{E}[e^X] \leq \mathbb{E}\left[\left(\frac{X+a}{2a}\right)e^a + \left(\frac{a-X}{2a}\right)e^{-a}\right]\,.$$

Using the linearity of expectation, we get:

$$\mathbb{E}[e^X] \leq \frac{\mathbb{E}[X] + a}{2a}e^a + \frac{a - \mathbb{E}[X]}{2a}e^{-a}\,.$$

We are given that $\mathbb{E}[X] = 0$, so substituting this into the inequality yields:

$$\mathbb{E}[e^X] \leq \frac{a}{2a}e^a + \frac{a}{2a}e^{-a} = \frac{e^a + e^{-a}}{2} = \cosh(a)\,.$$

This establishes that $\cosh(a)$ is an upper bound for $\mathbb{E}[e^X]$.

To show that this bound is the maximum, we must show it is achievable. Consider a discrete random variable X such that:

$$P(X = a) = \frac{1}{2} \quad \text{and} \quad P(X = -a) = \frac{1}{2}\,.$$

This variable satisfies the given conditions: it is bounded by $[-a, a]$, and its expectation is $\mathbb{E}[X] = a(\frac{1}{2}) + (-a)(\frac{1}{2}) = 0$. For this distribution, the expectation of e^X is:

$$\mathbb{E}[e^X] = e^a \cdot P(X = a) + e^{-a} \cdot P(X = -a) = \frac{1}{2}e^a + \frac{1}{2}e^{-a} = \cosh(a)\,.$$

Since the upper bound is achieved for this distribution, the maximum value of $\mathbb{E}[e^X]$ is $\cosh(a)$.

Problem 8 (Almost Sure Convergence of Scaled Random Walks):

Let X be a random variable with distribution $P(X = 1) = P(X = -1) = \frac{1}{2}$. Define the partial sum $S_n = X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.) copies of X. For any $\alpha > \frac{1}{2}$, prove that

$$P\left(\lim_{n\to\infty} \frac{S_n}{n^\alpha} = 0\right) = 1\,.$$

Solution 8 (Almost Sure Convergence of Scaled Random Walks):

The proof here is to show that for any arbitrarily small $\epsilon > 0$, the probability of $|Y_n|$ exceeding ϵ infinitely often is zero.

The elementary properties of the random variable X_i are

$$\mathbb{E}[X_i] = (1) \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$$

and

$$\mathsf{Var}(X_i) = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = (1^2 \cdot \frac{1}{2} + (-1)^2 \cdot \frac{1}{2}) - 0^2 = 1$$

Since the X_i are i.i.d., the expectation and variance of the partial sum S_n are:

$$\mathbb{E}[S_n] = \sum_{i=1}^{n} \mathbb{E}[X_i] = 0$$

and

$$\mathsf{Var}(S_n) = \sum_{i=1}^{n} \mathsf{Var}(X_i) = n$$

A common strategy for proving almost sure convergence is to apply the First Borel-Cantelli Lemma. To do this, we must show that for any $\epsilon > 0$, the sum of the probabilities of the events $\{|S_n/n^\alpha| > \epsilon\}$ is finite. We can bound these probabilities using Chebyshev's inequality:

$$P\left(\left|\frac{S_n}{n^\alpha}\right| > \epsilon\right) = P\left(|S_n| > \epsilon n^\alpha\right) \le \frac{\mathsf{Var}(S_n)}{(\epsilon n^\alpha)^2} = \frac{n}{\epsilon^2 n^{2\alpha}} = \frac{1}{\epsilon^2} n^{1-2\alpha}.$$

Now, we consider the sum of these probabilities:

$$\sum_{n=1}^{\infty} P\left(\left|\frac{S_n}{n^\alpha}\right| > \epsilon\right) \le \sum_{n=1}^{\infty} \frac{1}{\epsilon^2} n^{1-2\alpha} = \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^{2\alpha-1}}.$$

This is a p-series, which converges if and only if the exponent is greater than 1. Thus, we require $2\alpha - 1 > 1$, which simplifies to $\alpha > 1$. This approach, therefore, is insufficient to prove the claim for the entire range $\alpha > 1/2$, as it fails for the case $1/2 < \alpha \le 1$.

To provide a complete proof covering the entire range $\alpha > 1/2$, we invoke the Law of the Iterated Logarithm (LIL), a much stronger result that precisely characterizes the magnitude of the fluctuations of S_n. The LIL states that

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1, \quad \text{almost surely.}$$

This implies that for almost every sample path, the sequence $\frac{S_n}{\sqrt{2n \log\log n}}$ is bounded. Specifically, for almost every outcome ω, there exists a constant $C(\omega)$ such that for all sufficiently large n,

$$\left|\frac{S_n(\omega)}{\sqrt{2n \log\log n}}\right| \le C(\omega).$$

Now, let's examine the limit of the expression in our problem. We can rewrite it as:

$$\left|\frac{S_n}{n^\alpha}\right| = \left|\frac{S_n}{\sqrt{2n \log\log n}}\right| \cdot \frac{\sqrt{2n \log\log n}}{n^\alpha}.$$

We analyze the limit of this product as $n \to \infty$.

1. The first term, $\left|\frac{S_n}{\sqrt{2n \log \log n}}\right|$, is almost surely bounded due to the LIL.

2. The second term can be analyzed as follows:

$$\frac{\sqrt{2n \log \log n}}{n^\alpha} = \sqrt{2 \log \log n} \cdot \frac{n^{1/2}}{n^\alpha} = \sqrt{2 \log \log n} \cdot n^{1/2-\alpha}.$$

Since we are given that $\alpha > 1/2$, the exponent $1/2 - \alpha$ is strictly negative. A term with polynomial decay, n^{-p} for $p > 0$, goes to zero much faster than a term with logarithmic growth, $\sqrt{\log \log n}$. Therefore, the limit of the second term is zero:

$$\lim_{n\to\infty} \sqrt{2 \log \log n} \cdot n^{1/2-\alpha} = 0.$$

Since we are multiplying an almost surely bounded sequence (the first term) by a deterministic sequence that converges to zero (the second term), the overall product must converge to zero almost surely. Therefore,

$$\lim_{n\to\infty} \frac{S_n}{n^\alpha} = 0, \quad \text{almost surely.}$$

This establishes the desired result for all $\alpha > 1/2$.

Problem 9 (Probability Bound for the Standardized Sum of Uniform Random Variables):
Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables uniformly distributed on the interval $(-1, 1)$. For any $r > 0$, prove that

$$P\left(\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} < r\right) > 1 - \frac{1}{3r^2}.$$

Solution 9 (Probability Bound for the Standardized Sum of Uniform Random Variables):

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables, uniformly distributed on the interval $(-1, 1)$. We want to prove that for any $r > 0$,

$$P\left(\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} < r\right) > 1 - \frac{1}{3r^2}.$$

First, we determine the expectation and variance of each individual random variable X_i. Since each X_i follows a uniform distribution on the interval (a, b) with $a = -1$ and $b = 1$, we use the standard formulas for the expectation and variance of a uniform distribution.

The formula for the expectation of a uniform random variable on (a, b) is $\mathbb{E}[X] = \frac{a+b}{2}$. For $X_i \sim U(-1, 1)$, the expectation is:

$$\mathbb{E}[X_i] = \frac{-1+1}{2} = 0.$$

The formula for the variance of a uniform random variable on (a, b) is $\text{Var}(X) = \frac{(b-a)^2}{12}$. The variance of X_i is:

$$\text{Var}(X_i) = \frac{(1-(-1))^2}{12} = \frac{2^2}{12} = \frac{4}{12} = \frac{1}{3}.$$

Next, let's define the sum of these random variables as $S_n = X_1 + X_2 + \cdots + X_n$. We compute the expectation and variance of S_n.

By the linearity of expectation, the expectation of a sum is the sum of the expectations: $\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i]$. Thus, the expectation of S_n is:

$$\mathbb{E}[S_n] = \sum_{i=1}^n \mathbb{E}[X_i] = n \cdot 0 = 0.$$

Since the random variables X_i are independent, the variance of a sum of independent random variables is the sum of their variances: $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$. Therefore, the variance of S_n is:

$$\text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i) = n \cdot \frac{1}{3} = \frac{n}{3}.$$

The quantity we are interested in is the standardized sum, $\frac{S_n}{\sqrt{n}}$. We find its expectation and variance. The expectation is $\mathbb{E}\left[\frac{S_n}{\sqrt{n}}\right] = \frac{1}{\sqrt{n}}\mathbb{E}[S_n] = 0$. To find the variance, we use the property that for any random variable Y and constant c, $\text{Var}(cY) = c^2\text{Var}(Y)$.

$$\text{Var}\left(\frac{S_n}{\sqrt{n}}\right) = \text{Var}\left(\frac{1}{\sqrt{n}}S_n\right) = \left(\frac{1}{\sqrt{n}}\right)^2 \text{Var}(S_n) = \frac{1}{n} \cdot \frac{n}{3} = \frac{1}{3}.$$

Now we apply Chebyshev's inequality to the random variable $Z_n = \frac{S_n}{\sqrt{n}}$, which has mean $\mu = 0$ and variance $\sigma^2 = \frac{1}{3}$.

Chebyshev's inequality states that for a random variable Y with finite mean μ and finite non-zero variance σ^2, for any $k > 0$:

$$P(|Y - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Applying this inequality with $Y = \frac{S_n}{\sqrt{n}}$, $\mu = 0$, $\sigma^2 = \frac{1}{3}$, and setting $k = r$, we get:

$$P\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right) \leq \frac{1/3}{r^2} = \frac{1}{3r^2}\,.$$

Our goal is to find a lower bound for $P\left(\frac{S_n}{\sqrt{n}} < r\right)$. We can relate this to the quantity above using probability rules. First, we use the rule for complementary events, $P(A) = 1 - P(A^c)$. The complement of the event $\left\{\frac{S_n}{\sqrt{n}} < r\right\}$ is $\left\{\frac{S_n}{\sqrt{n}} \geq r\right\}$. So,

$$P\left(\frac{S_n}{\sqrt{n}} < r\right) = 1 - P\left(\frac{S_n}{\sqrt{n}} \geq r\right)\,.$$

Next, we observe the relationship between the events $\left\{\frac{S_n}{\sqrt{n}} \geq r\right\}$ and $\left\{\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right\}$. The event $\left\{\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right\}$ is the union of two disjoint events: $\left\{\frac{S_n}{\sqrt{n}} \geq r\right\} \cup \left\{\frac{S_n}{\sqrt{n}} \leq -r\right\}$. Therefore, the event $\left\{\frac{S_n}{\sqrt{n}} \geq r\right\}$ is a subset of the event $\left\{\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right\}$.

By the monotonicity of probability, if $A \subseteq B$, then $P(A) \leq P(B)$. This implies:

$$P\left(\frac{S_n}{\sqrt{n}} \geq r\right) \leq P\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right)\,.$$

This leads to $-P\left(\frac{S_n}{\sqrt{n}} \geq r\right) \geq -P\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right)$, and by adding 1 to both sides, we get:

$$1 - P\left(\frac{S_n}{\sqrt{n}} \geq r\right) \geq 1 - P\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right)\,.$$

Substituting back the complementary probability, we have:

$$P\left(\frac{S_n}{\sqrt{n}} < r\right) \geq 1 - P\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right)\,.$$

Finally, we combine this result with the bound from Chebyshev's inequality:

$$P\left(\frac{S_n}{\sqrt{n}} < r\right) \geq 1 - P\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right) \geq 1 - \frac{1}{3r^2}\,.$$

This establishes the non-strict inequality $P\left(\frac{S_n}{\sqrt{n}} < r\right) \geq 1 - \frac{1}{3r^2}$.

To justify the strict inequality required by the problem, we note that the distribution of S_n is continuous and its support is $(-n, n)$, so the support of $\frac{S_n}{\sqrt{n}}$ is $(-\sqrt{n}, \sqrt{n})$. Case 1: $r \geq \sqrt{n}$. The event $\frac{S_n}{\sqrt{n}} < r$ is certain to happen, so $P(\frac{S_n}{\sqrt{n}} < r) = 1$. The inequality to prove is $1 > 1 - \frac{1}{3r^2}$, which simplifies to $\frac{1}{3r^2} > 0$. This is true since $r > 0$. Case 2: $0 < r < \sqrt{n}$. In this case, the event $\left\{\frac{S_n}{\sqrt{n}} \leq -r\right\}$ has a probability strictly greater than zero, i.e., $P\left(\frac{S_n}{\sqrt{n}} \leq -r\right) > 0$. This implies that the inequality $P\left(\frac{S_n}{\sqrt{n}} \geq r\right) \leq P\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right)$ must be strict, because $P\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right) = P\left(\frac{S_n}{\sqrt{n}} \geq r\right) + P\left(\frac{S_n}{\sqrt{n}} \leq -r\right)$. Therefore, $P\left(\frac{S_n}{\sqrt{n}} \geq r\right) < P\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right)$, which makes our derivation yield a strict inequality:

$$P\left(\frac{S_n}{\sqrt{n}} < r\right) = 1 - P\left(\frac{S_n}{\sqrt{n}} \geq r\right) > 1 - P\left(\left|\frac{S_n}{\sqrt{n}}\right| \geq r\right) \geq 1 - \frac{1}{3r^2}$$

Thus, for any $r > 0$, the strict inequality holds:

$$P\left(\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} < r\right) > 1 - \frac{1}{3r^2}\,.$$

Problem 10 (Central Limit Theorem and Standardized Sum Convergence):

Let $\{X_1, X_2, \ldots, X_n\}$ be a sequence of independent and identically distributed (i.i.d.) random variables with finite mean $\mu = \mathbb{E}[X_i]$ and finite variance $\sigma^2 = \mathsf{Var}(X_i) > 0$. Define the standardized sum:

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then, as $n \to \infty$, the distribution of Z_n converges to the standard normal distribution $\mathcal{N}(0,1)$. Formally, for all real numbers z:

$$\lim_{n\to\infty} P(Z_n \le z) = \Phi(z),$$

where $\Phi(z)$ is the cumulative distribution function (CDF) of the standard normal distribution.

Solution 10 (Central Limit Theorem and Standardized Sum Convergence):

Let us define a new sequence of i.i.d. random variables $Y_i = X_i - \mu$. These are the centered versions of X_i. The moments of Y_i are:

$$\mathbb{E}[Y_i] = \mathbb{E}[X_i - \mu] = \mu - \mu = 0$$

$$\mathsf{Var}(Y_i) = \mathsf{Var}(X_i - \mu) = \mathsf{Var}(X_i) = \sigma^2$$

So, we have $\mathbb{E}[Y_i^2] = \mathsf{Var}(Y_i) + (\mathbb{E}[Y_i])^2 = \sigma^2$.

We can rewrite Z_n in terms of Y_i:

$$Z_n = \frac{\sum_{i=1}^n Y_i}{\sigma\sqrt{n}} = \sum_{i=1}^n \frac{Y_i}{\sigma\sqrt{n}}$$

Let $\phi_{Z_n}(t)$ be the characteristic function of Z_n. Since the terms $\frac{Y_i}{\sigma\sqrt{n}}$ are i.i.d., we can use the property that the characteristic function of a sum of independent random variables is the product of their individual characteristic functions.

$$\phi_{Z_n}(t) = \mathbb{E}\left[e^{itZ_n}\right] = \mathbb{E}\left[\exp\left(it\sum_{i=1}^n \frac{Y_i}{\sigma\sqrt{n}}\right)\right] = \prod_{i=1}^n \mathbb{E}\left[\exp\left(it\frac{Y_i}{\sigma\sqrt{n}}\right)\right]$$

As the Y_i are identically distributed, this simplifies to:

$$\phi_{Z_n}(t) = \left[\phi_{Y/(\sigma\sqrt{n})}(t)\right]^n$$

where $\phi_{Y/(\sigma\sqrt{n})}(t)$ is the characteristic function of any of the variables $\frac{Y_i}{\sigma\sqrt{n}}$.

We use the scaling property of characteristic functions: $\phi_{aX}(t) = \phi_X(at)$ for any constant a. Let $s = \frac{t}{\sigma\sqrt{n}}$. Then:

$$\phi_{Z_n}(t) = \left[\phi_Y\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

Now, we find the Taylor expansion of the characteristic function $\phi_Y(s)$ around $s = 0$. If a random variable Y has finite moments up to order k, its characteristic function $\phi_Y(s)$ can be expanded as:

$$\phi_Y(s) = \sum_{j=0}^k \frac{\mathbb{E}[Y^j]}{j!}(is)^j + o(s^k)$$

Since $\mathbb{E}[Y] = 0$ and $\mathbb{E}[Y^2] = \sigma^2$, we expand up to the second order ($k = 2$):

$$\phi_Y(s) = \frac{\mathbb{E}[Y^0]}{0!}(is)^0 + \frac{\mathbb{E}[Y^1]}{1!}(is)^1 + \frac{\mathbb{E}[Y^2]}{2!}(is)^2 + o(s^2)$$

$$\phi_Y(s) = 1 + 0 + \frac{\sigma^2}{2}(-s^2) + o(s^2) = 1 - \frac{\sigma^2 s^2}{2} + o(s^2)$$

Substitute $s = \frac{t}{\sigma\sqrt{n}}$ into this expansion. As $n \to \infty$, $s \to 0$.

$$\phi_Y\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)$$

$$\phi_Y\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{\sigma^2 t^2}{2\sigma^2 n} + o\left(\frac{t^2}{n}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)$$

Now, we substitute this back into the expression for $\phi_{Z_n}(t)$:

$$\phi_{Z_n}(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n$$

To find the limit as $n \to \infty$, we use the well-known result from calculus: $\lim_{n\to\infty}\left(1 + \frac{a}{n} + o\left(\frac{1}{n}\right)\right)^n = e^a$. With $a = -t^2/2$, we get:

$$\lim_{n\to\infty} \phi_{Z_n}(t) = e^{-t^2/2}$$

This resulting function, $\phi(t) = e^{-t^2/2}$, is the characteristic function of the standard normal distribution, $\mathcal{N}(0,1)$.

By Levy's Continuity Theorem, since the characteristic functions $\phi_{Z_n}(t)$ converge pointwise to the characteristic function of a standard normal distribution, the distribution of Z_n converges to the standard normal distribution. That is:

$$Z_n \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as } n \to \infty.$$

This completes the proof.