Short HW1

Murad Ektilat

June 13, 2024

1 Variance of empirical mean

1.1

Given $\bar{X} = \frac{1}{m}\sum_i X_i$ and $E(X_i) = \mu$:

$E(\bar{X}) \underbrace{=}_{\text{def of } \bar{X}} E(\frac{1}{m}\sum_i X_i) \underbrace{=}_{\text{linearity of E}} \frac{1}{m}\sum_i E(X_i) = \frac{1}{m}\sum_i \mu = \frac{1}{m} m\mu = \mu$

1.2

Given $var(X_i) = \sigma^2$:

$var(\bar{X}) \underbrace{=}_{\text{def of } \bar{X}} var(\frac{1}{m}\sum_i X_i) = (\frac{1}{m})^2 var(\sum_i X_i) \underbrace{=}_{\text{i.i.d}} \frac{1}{m^2}\sum_i var(X_i) = \frac{1}{m^2}\sum_i \sigma^2$
$= \frac{1}{m^2} m\sigma^2 = \frac{\sigma^2}{m}$

1.3

code

2 Identical coins and the Hoeffding bound

2.1

θ_i is a random variable describing the "success rate" (getting heads) of a series of binary trials (coin tossing) with fixed probability p.
Hence $\forall i : \theta_i \sim$ Bin(40,p).

2.2

The expexted value of a random variable with Binomial distibution is np, hence $\forall i : E(\theta_i) = 40p$

2.3

According to Hoeffding's Inequality, $P[|\bar{\theta}(m) - \mu| > \epsilon] \leq 2exp(-\frac{2m\epsilon^2}{(b-a)^2})$.

To achive $P[|\bar{\theta}(m) - \mu| > \epsilon] \leq 0.05$ we need to find ϵ s.t. $2exp(-\frac{2m\epsilon^2}{(b-a)^2}) = 0.05$

plugging in $m = 20, a = 0, b = 40$: we get:

$2exp(-\frac{2*20\epsilon^2}{40^2}) = 0.05$

$exp(-\frac{\epsilon^2}{40}) = 0.025$

$-\frac{\epsilon^2}{40} = ln(0.025)$

$\epsilon = 12.147$

2.4

code

2.5

The estimators mean has a normal distribution. The central limit theorem explains this behavior. The theorem applies because the sample size is large enough ($tosses = 50 > 30$), the data points in the sample are independent of each other, and the variance is finite.

3 PSD matrices

3.1

$A\vec{v} = \lambda\vec{v}$ / multiply by $\vec{v}^T$ from the left

$\vec{v}^T A\vec{v} = \vec{v}^T\lambda\vec{v}$ / A is PSD, hence $\forall\vec{v} : \vec{v}^T A\vec{v} \geq 0$

$\vec{v}^T\lambda\vec{v} \geq 0$ / λ is scalar

$\lambda\vec{v}^T\vec{v} \geq 0$ / $\vec{v}^T = (v_1, ...v_n)$

$\lambda\sum\limits_i v_i^2 \geq 0$ / $\vec{v} \neq \vec{0}, \forall i : v_i^2 \geq 0 \Rightarrow \sum\limits_i v_i^2 > 0$

$\lambda \geq 0$

3.2

$B^T AB$ is PSD. proof:

- $B^T AB$ is symmetric:

  $(B^T AB)^T = B^T A^T (B^T)^T \underbrace{=}_{\text{A is symmetric}} B^T AB$

- $\vec{v}^T(B^T AB)\vec{v} = (B\vec{v})^T AB\vec{v} \underbrace{=}_{\vec{x}=B\vec{v}} \vec{x}^T A\vec{x} \underbrace{\geq}_{\text{A is PSD}} 0$

For $B^T AB$ to be PD we need $\vec{v}^T(B^T AB)\vec{v} > 0$

Looking at the last Inequality we got: $\vec{x}^T A\vec{x} \geq 0$ we infere two conditions:

1. A must be PD

2. B must be non-invertible

($\vec{x}$ must be a non zero vector. As $\vec{x} = B\vec{v}$ and $\vec{v}$ could be any non-zero vector, we must demand that B will be non-invertible.)

4 Gradients

4.1

$f(\vec{w}) = \vec{w}^T\vec{x} + b = \sum_i (w_i x_i) + b$

$\boldsymbol{\nabla}_w f(\vec{w}) = (\frac{\partial f(\vec{w})}{\partial w_1}, ..., \frac{\partial f(\vec{w})}{\partial w_d}) \underbrace{=}_{\frac{\partial f(\vec{w})}{\partial w_i} = x_i} (x_1, ..., x_d) = \vec{x}$

4.2

$\boldsymbol{\nabla}^2_w f(\vec{w})_{m,n} = \frac{\partial^2 f(\vec{w})}{\partial w_m \partial w_n}$

$\forall m, n : x_m$ is independent of $w_n \Rightarrow \boldsymbol{\nabla}^2_w f(\vec{w})_{m,n} = 0 \Rightarrow \boldsymbol{\nabla}^2_w f(\vec{w}) = 0_{d\times d}$

4.3

$0_{d\times d}$ is a PSD as:

$\forall \vec{v} \neq \vec{0} : \vec{v}^T 0_{d\times d} \vec{v} = 0 \geq 0$

4.4

$g(\vec{w}) = \frac{1}{2}\lambda ||\vec{w}||^2 = \frac{1}{2}\lambda \sum_i w_i^2$

$\boldsymbol{\nabla}_w g(\vec{w}) = (\frac{1}{2}\lambda \cdot 2w_1, ..., \frac{1}{2}\lambda \cdot 2w_d) = \lambda(w_1, ..., w_d) = \lambda\vec{w}$

4.5

$$\boldsymbol{\nabla}^2_w g(\vec{w})_{m,n} = \frac{\partial^2 g(\vec{w})}{\partial w_m \partial w_n} = \begin{cases} 0, & m \neq n \\ \lambda, & m = n \end{cases} \Rightarrow \boldsymbol{\nabla}^2_w g(\vec{w}) = \lambda I_{d\times d}$$

4.6

$\lambda I_{d\times d}$ is a PD as:

$\forall \vec{v}\neq\vec{0}: \vec{v}^T\lambda I_{d\times d}\vec{v}=\lambda\vec{v}^T\vec{v}=\lambda\sum_i v_i^2 \underbrace{>}_{\lambda>0,\sum_i v_i^2>0} 0$

4.7

$h(w_1,w_2)=12w_1^3-36w_1w_2-2w_2^3+9w_2^2-72w_1+60w_2+5$

$\boldsymbol{\nabla}_{\vec{w}}h(\vec{w})=(36w_1^2-36w_2-72,-36w_1-6w_2^2+18w_2+60)$

$$\boldsymbol{\nabla}_{\vec{w}}h(\vec{w})=0\Rightarrow\begin{cases}36w_1^2-36w_2-72=0\\ -36w_1-6w_2^2+18w_2+60=0\end{cases}\Rightarrow\begin{cases}w_2=w_1^2-2\\ w_2^2+6w_1-3w_2-10=0\end{cases}\underbrace{\Rightarrow}_{\text{plugging } w_2}$$

$(w_1^2-2)^2+6w_1-3(w_1^2-2)-10=0$

$w_1^4-4w_1^2+4+6w_1-3w_1^2+6-10=0$

$w_1(w_1^3-7w_1+6)=0$

$w_1(w_1+3)(w_1-2)(w_1-1)=0$

$\vec{w}^*=(0,-2),(-3,7),(2,2),(1,-1)$

4.8

To determine max/min/saddle of the critical points, we will use the second partial derivative test. $h(w_1.w_2)$ is a polinomial real function, hence it is differentiable, and its second partial derivatives exist and are continuous.

$\boldsymbol{\nabla}^2_{\vec{w}}h(\vec{w})_{1,1}=\frac{\partial^2 h(\vec{w})}{\partial^2 w_1}=72w_1$

$\boldsymbol{\nabla}^2_{\vec{w}}h(\vec{w})_{1,2}=\frac{\partial^2 h(\vec{w})}{\partial w_1\partial w_2}=-36$

$\boldsymbol{\nabla}^2_{\vec{w}}h(\vec{w})_{2,1}=\frac{\partial^2 h(\vec{w})}{\partial w_2\partial w_1}=-36$

$\boldsymbol{\nabla}^2_{\vec{w}}h(\vec{w})_{2,2}=\frac{\partial^2 h(\vec{w})}{\partial^2 w_2}=-12w_2+18$

$|\boldsymbol{\nabla}^2_{\vec{w}}h(\vec{w})|=72w_1\cdot(-12w_2+18)-(-36)\cdot(-36)=-864w_1w_2+1296w_1-1296$

caculating the det and second derivative at each critical point we get:

$|\nabla^2_{\vec{w}} h(0,-2)| = -1296 < 0 \Rightarrow$ saddle point.

$|\nabla^2_{\vec{w}} h(2,2)| = -2160 < 0 \Rightarrow$ saddle point.

$|\nabla^2_{\vec{w}} h(-3,7)| = 12,960 > 0$ and $\frac{\partial^2 h(\vec{w})}{\partial^2 w_1} = -216 < 0 \Rightarrow$ local Max.

$|\nabla^2_{\vec{w}} h(1,-1)| = 864 > 0$ and $\frac{\partial^2 h(\vec{w})}{\partial^2 w_2} = 30 > 0 \Rightarrow$ local Min.

4.9

$h(w_1, w_2)$ is not bounded from below or from above, hence it has no global min/max. proof:

$h(w_1, 0) = 12w_1^3 - 72w_1 + 5$

$\lim_{w_1 \to \infty} h(w_1, 0) = \infty$

$h(0, w_2) = -2w_2^3 + 9w_2^2 + 5$

$\lim_{w_2 \to \infty} h(0, w_2) = -\infty$