
Covariance Sensitivity Proofs

March 22, 2020

1 PRELIMINARIES

Definition 1. Let X be a matrix of values and X_i indicate the i^{th} column of the matrix. Denote the sample mean of column X_i as \bar{X}_i , and let n be the size of X_i . Then the covariance matrix of X has ij^{th} element

$$\frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{X}_i)(x_{kj} - \bar{X}_j).$$

Lemma 1. Let X be a matrix of values and X_i indicate the i^{th} column of the matrix. Denote the sample mean of column X_i as \bar{X}_i , and let n be the size of X_i . Then, $\forall i$,

$$\sum_{j=1}^n (x_{ji} - \bar{X}_i) = 0.$$

Proof.

$$\begin{aligned} \sum_{j=1}^n (x_{ji} - \bar{X}_i) &= \sum_{j=1}^n x_{ji} - n\bar{X}_i, \\ &= \sum_{j=1}^n x_{ji} - n \left(\frac{1}{n} \sum_{j=1}^n x_{ji} \right), \\ &= 0. \end{aligned}$$

□

Lemma 2. Let X be a matrix and let

$$f_{ij}(X) = \sum_{k=1}^n (x_{ki} - \bar{X}_i)(x_{kj} - \bar{X}_j).$$

Note that this is equivalent to the ij^{th} element of the sample covariance matrix for X , without the normalization by $n-1$. Consider the matrix X' equal to X with a single row

Y added, so that $X'_i = X_i \cup \{y_i\}$. Say X_i has size n . Let \bar{X}_i , \bar{X}_j , \bar{X}'_i , and \bar{X}'_j be the sample means of X_i , X_j , X'_i and X'_j respectively. Then,

$$f_{ij}(X') = f_{ij}(X) + n(\bar{X}_i - \bar{X}'_i)(\bar{X}_j - \bar{X}'_j) + (y_i - \bar{X}_i)(y_j - \bar{X}_j).$$

Proof. Note that

$$\begin{aligned} f_{ij}(X') &= \sum_{k=1}^{n+1} (x'_{ki} - \bar{X}'_i)(x'_{kj} - \bar{X}'_j), \\ &= \sum_{k=1}^n (x_{ki} - \bar{X}_i)(x_{kj} - \bar{X}_j) + (y_i - \bar{X}_i)(y_j - \bar{X}_j), \\ &= \sum_{k=1}^n ((x_{ki} - \bar{X}_i) + (\bar{X}_i - \bar{X}'_i)) ((x_{kj} - \bar{X}_j) + (\bar{X}_j - \bar{X}'_j)) + (y_i - \bar{X}_i)(y_j - \bar{X}_j), \\ &= \sum_{k=1}^n (x_{ki} - \bar{X}_i)(x_{kj} - \bar{X}_j) + (\bar{X}_j - \bar{X}'_j) \sum_{k=1}^n (x_{ki} - \bar{X}_i) + (\bar{X}_i - \bar{X}'_i) \sum_{k=1}^n (x_{kj} - \bar{X}_j), \\ &\quad + \sum_{k=1}^n (\bar{X}_i - \bar{X}'_i)(\bar{X}_j - \bar{X}'_j) + (y_i - \bar{X}_i)(y_j - \bar{X}_j), \\ &= f_{ij}(X) + n(\bar{X}_i - \bar{X}'_i)(\bar{X}_j - \bar{X}'_j) + (y_i - \bar{X}_i)(y_j - \bar{X}_j), \end{aligned}$$

where the cancellation of the second and third terms in the second-to-last line is due to Lemma 1. \square

Lemma 3. Let X be a matrix of values and X_i indicate the i^{th} column of the matrix. Let X_i have size n and consider the matrix X' equal to X with a single row Y added, so that $X'_i = X_i \cup \{y_i\}$. Say that the space of datapoints \mathcal{X}_i that the elements of X'_i are drawn from is bounded above by M_i and bounded below by m_i . Let \bar{X}_i , \bar{X}_j , \bar{X}'_i , and \bar{X}'_j be the sample means of X_i , X_j , X'_i and X'_j respectively. Then,

$$n |(\bar{X}_i - \bar{X}'_i)(\bar{X}_j - \bar{X}'_j)| \leq \frac{n}{(n+1)^2} (M_i - m_i)(M_j - m_j).$$

Proof. Note that

$$\begin{aligned} n |(\bar{X}_i - \bar{X}'_i)(\bar{X}_j - \bar{X}'_j)| &= n \left| \left(\frac{1}{n} \sum_{k=1}^n x_{ki} - \frac{1}{n+1} \sum_{k=1}^{n+1} x'_{ki} \right) \left(\frac{1}{n} \sum_{k=1}^n x_{kj} - \frac{1}{n+1} \sum_{k=1}^{n+1} x'_{kj} \right) \right|, \\ &= n \left| \left(\left(\frac{1}{n} - \frac{1}{n+1} \right) \sum_{k=1}^n x_{ki} - \frac{y_i}{n+1} \right) \left(\left(\frac{1}{n} - \frac{1}{n+1} \right) \sum_{k=1}^n x_{kj} - \frac{y_j}{n+1} \right) \right|, \\ &= n \left| \left(\frac{1}{n(n+1)} \sum_{k=1}^n x_{ki} - \frac{y_i}{n+1} \right) \left(\frac{1}{n(n+1)} \sum_{k=1}^n x_{kj} - \frac{y_j}{n+1} \right) \right|, \\ &= \frac{n}{(n+1)^2} \left| \left(\frac{1}{n} \sum_{k=1}^n x_{ki} - \frac{y_i}{n+1} \right) \left(\frac{1}{n} \sum_{k=1}^n x_{kj} - \frac{y_j}{n+1} \right) \right|, \\ &\leq \frac{n}{(n+1)^2} (M_i - m_i)(M_j - m_j). \end{aligned}$$

\square

Lemma 4. Let X be a matrix of values and X_i indicate the i^{th} column of the matrix. Let X_i have size n and consider the matrix X' equal to X with a single row Y added, so that $X'_i = X_i \cup \{y_i\}$. Say that the space of datapoints \mathcal{X}_i that the elements of X'_i are drawn from is bounded above by M_i and bounded below by m_i . Let \bar{X}_i , \bar{X}_j , \bar{X}'_i , and \bar{X}'_j be the sample means of X_i , X_j , X'_i and X'_j respectively. Then,

$$|(y_i - \bar{X}'_i)(y_j - \bar{X}'_j)| \leq \frac{n^2}{(n+1)^2} (M_i - m_i)(M_j - m_j).$$

Proof. Note that

$$\begin{aligned} |(y_i - \bar{X}'_i)(y_j - \bar{X}'_j)| &= \left| \left(y_i - \frac{y_i + n\bar{X}_i}{n+1} \right) \left(y_j - \frac{y_j + n\bar{X}_j}{n+1} \right) \right|, \\ &= \frac{1}{(n+1)^2} |((n+1)y_i - y_i - n\bar{X}_i)((n+1)y_j - y_j - n\bar{X}_j)|, \\ &= \frac{n^2}{(n+1)^2} |(y_i - \bar{X}_i)(y_j - \bar{X}_j)|, \\ &\leq \frac{n^2}{(n+1)^2} (M_i - m_i)(M_j - m_j). \end{aligned}$$

□

2 NEIGHBORING DEFINITION: ADD/DROP ONE

2.1 ℓ_1 -sensitivity

Theorem 1. Let X be a matrix of values and let X_i indicate the i^{th} column of the matrix. Let

$$f_{ij}(X) = \sum_{k=1}^n (x_{ki} - \bar{X}_i)(x_{kj} - \bar{X}_j).$$

Say that the space of datapoints \mathcal{X}_i that X_i is drawn from is bounded above by M_i and bounded below by m_i . Then the ℓ_1 -sensitivity in the add/drop-one model of $f(\cdot)$ is bounded above by

$$\frac{n}{(n+1)} (M_i - m_i)(M_j - m_j).$$

Proof. We must consider both adding and removing a row from X .

Adding a row:

Let $X'_i = X_i \cup \{y_i\}$. Then, from Lemma 2,

$$\begin{aligned} |f_{ij}(X') - f_{ij}(X)| &= |n(\bar{X}_i - \bar{X}'_i)(\bar{X}_j - \bar{X}'_j) + (y_i - \bar{X}_i)(y_j - \bar{X}_j)| \\ &\leq n |(\bar{X}_i - \bar{X}'_i)(\bar{X}_j - \bar{X}'_j)| + |(y_i - \bar{X}_i)(y_j - \bar{X}_j)| \\ &\leq \frac{n}{(n+1)^2} (M_i - m_i)(M_j - m_j) + \frac{n^2}{(n+1)^2} (M_i - m_i)(M_j - m_j) \\ &\quad \text{(By Lemmas 3 and 4)} \\ &= \frac{n}{n+1} (M_i - m_i)(M_j - m_j). \end{aligned} \tag{2.1}$$

Removing a row:

Let Y be the last row of X , and let $X'_i = X_i \setminus \{y_i\}$. Note that Lemma 2 can be rewritten in this setting by parametrizing n as $n-1$ and swapping X and X' in its expression:

$$f_{ij}(X) = f_{ij}(X') + (n-1)(\bar{X}'_i - \bar{X}_i)(\bar{X}'_j - \bar{X}_j) + (y_i - \bar{X}'_i)(y_j - \bar{X}'_j).$$

Lemmas 3 and 4 may be rewritten with the same reparametrization:

$$(n-1)|(\bar{X}'_i - \bar{X}_i)(\bar{X}'_j - \bar{X}_j)| \leq \frac{n-1}{n^2}(M_i - m_i)(M_j - m_j),$$

and

$$|(y_i - \bar{X}'_i)(y_j - \bar{X}'_j)| \leq \frac{(n-1)^2}{n^2}(M_i - m_i)(M_j - m_j).$$

Then,

$$\begin{aligned} |f_{ij}(X) - f_{ij}(X')| &= |(n-1)(\bar{X}'_i - \bar{X}_i)(\bar{X}'_j - \bar{X}_j) + (y_i - \bar{X}'_i)(y_j - \bar{X}'_j)|, \\ &\leq |(n-1)(\bar{X}'_i - \bar{X}_i)(\bar{X}'_j - \bar{X}_j)| + |(y_i - \bar{X}'_i)(y_j - \bar{X}'_j)|, \\ &\leq \frac{n-1}{n^2}(M_i - m_i)(M_j - m_j) + \frac{(n-1)^2}{n^2}(M_i - m_i)(M_j - m_j), \\ &= \frac{n-1}{n}(M_i - m_i)(M_j - m_j). \end{aligned} \tag{2.2}$$

Note that for any $n \geq 1$,

$$\frac{n}{n+1} > \frac{n-1}{n}. \tag{2.3}$$

So, the worst-case bound always occurs in the “add-one” case, and in general the ℓ_1 sensitivity of $f(\cdot)$ is bounded by

$$\frac{n}{n+1}(M_i - m_i)(M_j - m_j).$$

□

Corollary 1. *Let $X \leftarrow \mathcal{X}$ where \mathcal{X}_i is bounded above by M_i and bounded below by m_i . Then the ℓ_1 -sensitivity in the add/drop-one model of the ij^{th} element of a sample covariance matrix for X is bounded above by*

$$\frac{n}{n^2 - 1}(M_i - m_i)(M_j - m_j).$$

Proof. Note that the sample covariance of X is equal to $f(x)/(n-1)$, and that $(n-1)(n+1) = n^2 - 1$. □

2.2 ℓ_2 -sensitivity

Theorem 2. *Let $X \leftarrow \mathcal{X}$ where \mathcal{X}_i is bounded above by M_i and bounded below by m_i . Then the ℓ_2 -sensitivity in the add/drop-one model of the ij^{th} element of a sample covariance matrix for X is bounded above by*

$$\left(\frac{n}{n^2 - 1}(M_i - m_i)(M_j - m_j) \right)^2.$$

Proof. This follows from the bounds in Equations 2.1 and 2.2 and the inequality in Equation 2.3. □

3 NEIGHBORING DEFINITION: CHANGE ONE

3.1 ℓ_1 -sensitivity

Theorem 3. *Let X be a matrix of values and let X_i indicate the i^{th} column of the matrix. Let*

$$f_{ij}(X) = \sum_{k=1}^n (x_{ki} - \bar{X}_i)(x_{kj} - \bar{X}_j).$$

Say that the space of datapoints \mathcal{X}_i that X_i is drawn from is bounded above by M_i and bounded below by m_i . Then the ℓ_1 -sensitivity in the change-one model of $f(\cdot)$ is bounded above by

$$\frac{2(n-1)}{n}(M_i - m_i)(M_j - m_j).$$

Proof. Recall from Equation 2.1 that

$$|f_{ij}(X) - f_{ij}(X')| \leq \frac{n}{(n+1)}(M_i - m_i)(M_j - m_j).$$

and

$$|f_{ij}(X) - f_{ij}(X'')| \leq \frac{n}{(n+1)}(M_i - m_i)(M_j - m_j).$$

Reparametrizing these equations so that n is the size of X' and X'' gives that

$$|f_{ij}(X) - f_{ij}(X')| \leq \frac{n-1}{n}(M_i - m_i)(M_j - m_j).$$

and

$$|f_{ij}(X) - f_{ij}(X'')| \leq \frac{n-1}{n}(M_i - m_i)(M_j - m_j).$$

It then follows from the triangle inequality that

$$|f_{ij}(X') - f_{ij}(X'')| \leq \frac{2(n-1)}{n}(M_i - m_i)(M_j - m_j).$$

3.2 ℓ_2 -sensitivity

□

Corollary 2. *The ℓ_1 -sensitivity in the change-one model of sample covariance is bounded above by*

$$\frac{2}{n}(M_i - m_i)(M_j - m_j).$$

Proof. Note that the sample covariance of X is equal to $f(x)/(n-1)$. □

3.3 ℓ_2 -sensitivity

Theorem 4. *The ℓ_2 -sensitivity in the change-one model of sample covariance is bounded above by*

$$\left(\frac{2}{n}(M_i - m_i)(M_j - m_j) \right)^2.$$

Proof. This follows from the bounds in the proof of Theorem 3. □