Covariance Sensitivity Proofs

March 21, 2020

1 PRELIMINARIES

Lemma 1. $\forall i$,

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

Proof.

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

$\square$

Lemma 2. *Let*

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

and let $X_i' = X_i \cup \{y_i\}$, and say X_i has size n. Let $\bar{X}_i'$ and $\bar{X}_j'$ be the sample means of 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

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

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

Lemma 3. *Let X_i have size n and say $X'_i = X_i \cup \{y_i\}$ where $y_i \in \mathcal{X}_i$. Say that the space of datapoints $\mathcal{X}_i$ is bounded above by M_i and bounded below by m_i, and let $y_i \in \mathcal{X}_i$. 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,*

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

Proof. Note that

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

□

Lemma 4. *Let X_i have size n and say $X'_i = X_i \cup \{y_i\}$ where $y_i \in \mathcal{X}_i$. Say that the space of datapoints $\mathcal{X}_i$ is bounded above by M_i and bounded below by m_i, and let $y_i \in \mathcal{X}_i$. 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,*

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

Proof. Note that

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

□

2 Neighboring Definition: Change One

2.1 ℓ_1-sensitivity

2.2 ℓ_2-sensitivity

3 Neighboring Definition: Add/Drop One

3.1 ℓ_1-sensitivity

3.2 ℓ_2-sensitivity