Automatic Classifiers as Scientific Instruments: One Step Further Away from Ground-Truth

Jacob Whitehill and Anand Ramakrishnan
Worcester Polytechnic Institute (WPI), Massachusetts, USA

ICML'2019

WPI

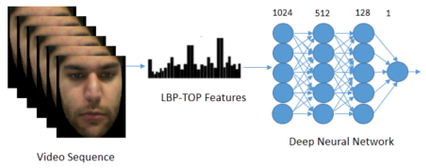
Machine learning to advance basic science

- Machine perception can advance basic science in:
  - Psychology
  - Education
  - Medicine

- …by providing automatic classifiers as new scientific instruments, e.g.:
  - Automatic stress detectors from wrist monitors instead of questionnaires.
  - Facial action unit detectors from video instead of electromyography.
  - Student engagement detectors from video instead of observational protocols.

Empatica E4 EDA

Emotient/ iMotions



Kaur et al. 2018

Correlation study

- Suppose a researcher wishes to measure the relationship between two constructs *U* and *V*, e.g.:
  - *U* = stress
  - *V* = academic performance.
- Standard methodology:
  - Use a **standard measurement tool** (e.g., survey, observational protocol) to estimate the values of *U* and *V* from a sample of n participants.
  - This produces two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, which we can assume w.l.o.g. have 0-mean and 1-length.
  - Estimate the correlation between *U* and *V* as:

$$r = \rho(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\top \mathbf{v} = \cos \angle(\mathbf{u}, \mathbf{v})$$

Only the angle between the two vectors determines their correlation.

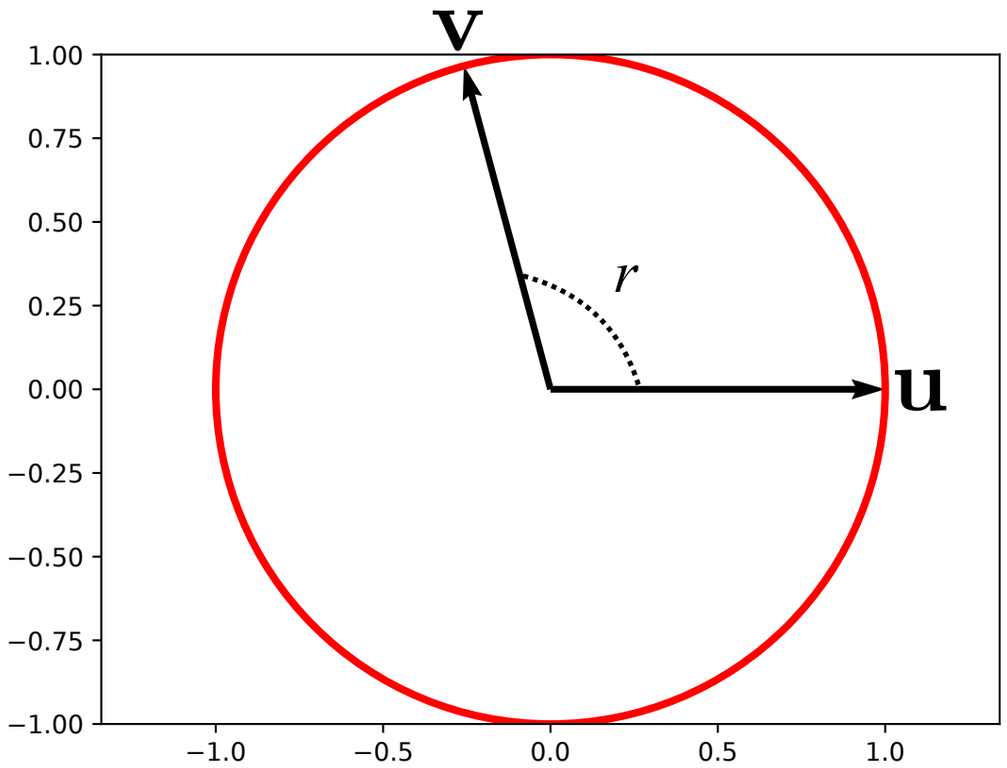
Correlation study

- But what if the researcher instead uses an **automatic stress detector** d whose correlation with ground-truth measurements is q (known from prior validation)?

- Instead of $\mathbf{u}$, the researcher obtains a vector $\hat{\mathbf{u}}$.

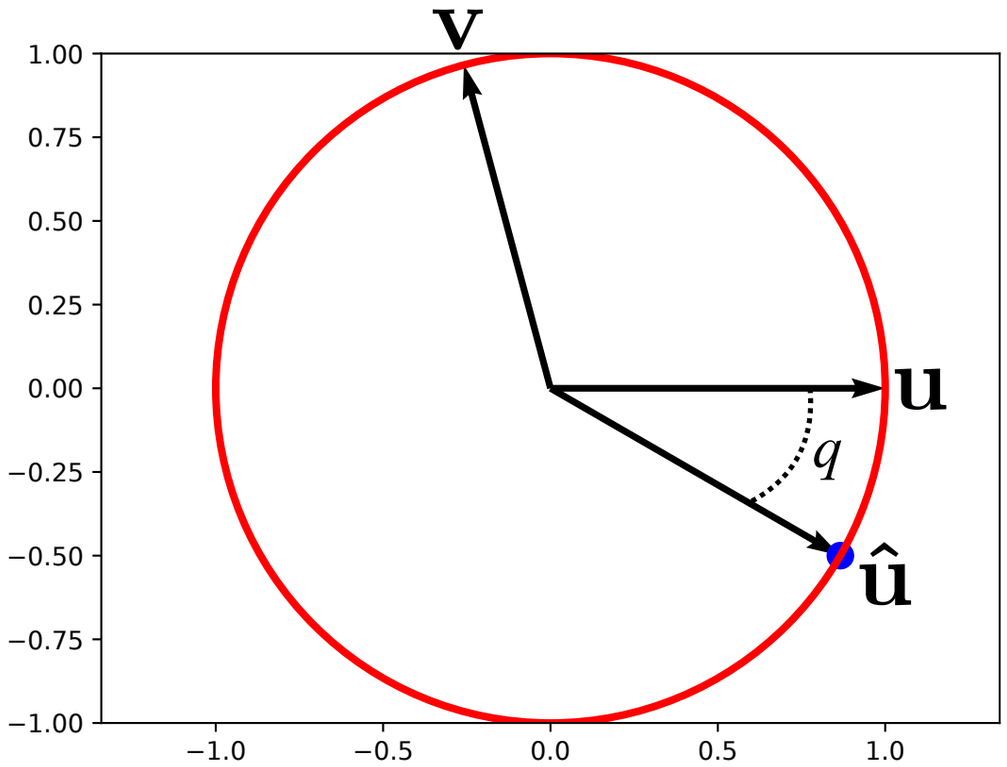
- What kind of spurious deductions about the correlation between *U* and *V* could result?

Trivariate correlation



- Suppose $\mathbf{u}$ and $\mathbf{v}$ are ground-truth values of *U* and *V*.
  - The correlation between $\mathbf{u}$ and $\mathbf{v}$ is $r = \cos(105°) = -.259$.

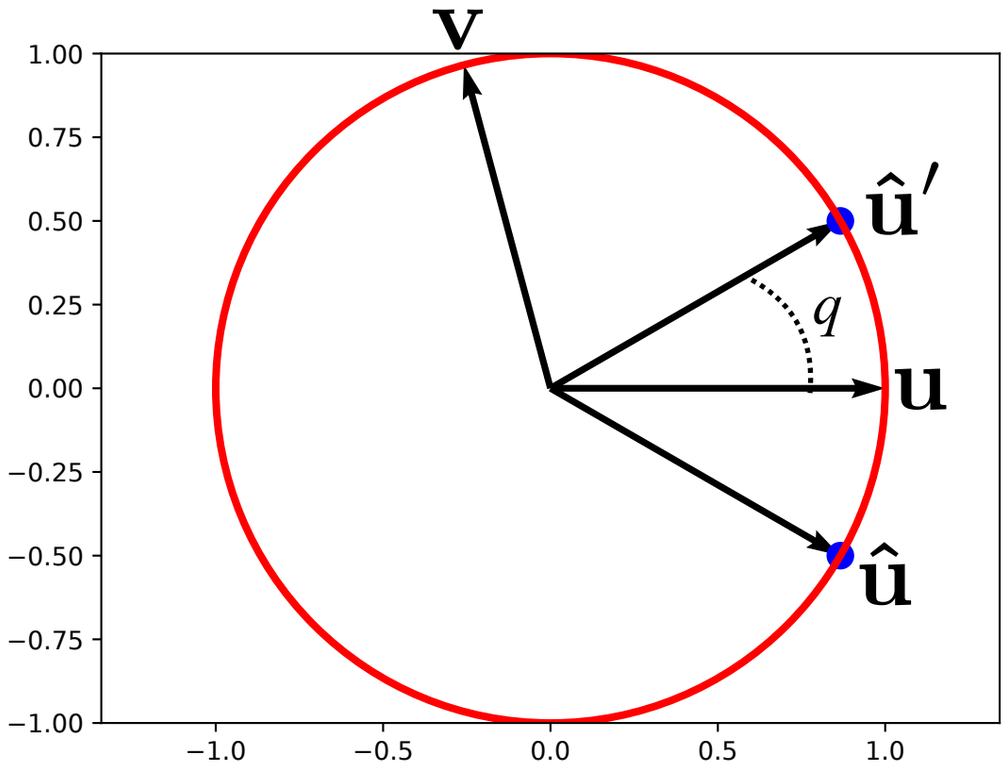
Trivariate correlation



- Using a detector d, the researcher might obtain $\hat{\mathbf{u}}$, whose correlation with $\mathbf{u}$ is q.

  - The correlation between $\hat{\mathbf{u}}$ and $\mathbf{v}$ is cos(135°)= -.707 — much larger than, but same sign as, the ground-truth correlation.

Trivariate correlation



- But they might also obtain vector $\hat{\mathbf{u}}'$, whose correlation with $\mathbf{u}$ is also q.

  - The correlation between $\hat{\mathbf{u}}'$ and $\mathbf{v}$ is cos(75°)= +.259 — this is the **opposite sign** as the ground-truth correlation.

We call this a *false correlation.*

Main results

1. The set of all vectors whose correlation with $\mathbf{u}$ is q, is an (n-3)-sphere $\mathcal{T}^n \in \mathbb{R}^n$.

Main results

1. The set of all vectors whose correlation with $\mathbf{u}$ is q, is an (n-3)-sphere $\mathcal{T}^n \in \mathbb{R}^n$.

2. If the correlation between $\mathbf{u}$ and $\mathbf{v}$ is r, then the expected sample correlation between $\hat{\mathbf{u}}$ and $\mathbf{v}$, where $\hat{\mathbf{u}}$ is drawn uniformly at random from $\mathcal{T}^n$, is qr.

Main results

1. The set of all vectors whose correlation with $\mathbf{u}$ is q, is an (n-3)-sphere $\mathcal{T}^n \in \mathbb{R}^n$.

2. If the correlation between $\mathbf{u}$ and $\mathbf{v}$ is r, then the expected sample correlation between $\hat{\mathbf{u}}$ and $\mathbf{v}$, where $\hat{\mathbf{u}}$ is drawn uniformly at random from $\mathcal{T}^n$, is qr.

3. We derive a formula $h(n,q,r)$ for the probability of a false correlation.

Main results

1. The set of all vectors whose correlation with $\mathbf{u}$ is q, is an (n-3)-sphere $\mathcal{T}^n \in \mathbb{R}^n$.

2. If the correlation between $\mathbf{u}$ and $\mathbf{v}$ is r, then the expected sample correlation between $\hat{\mathbf{u}}$ and $\mathbf{v}$, where $\hat{\mathbf{u}}$ is drawn uniformly at random from $\mathcal{T}^n$, is qr.

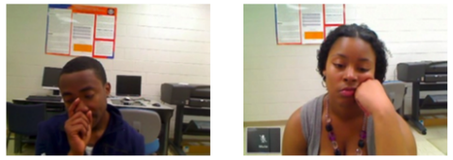
3. We derive a formula $h(n,q,r)$ for the probability of a false correlation.

4. We show that h is monotonically decreasing in q and n.

   But it can still be non-negligible for values of n, q used in recent affective computing studies — despite a small p-value.

Case study: Student engagement vs. cognitive task performance

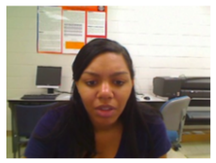
_U_: Engagement



(a) Engagement = 1

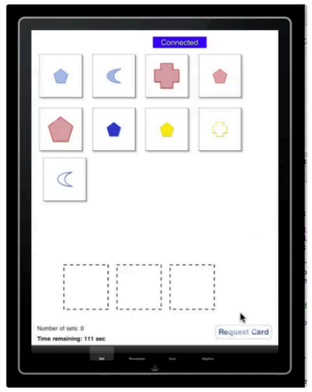
(b) Engagement = 2

(c) Engagement = 3



(d) Engagement = 4

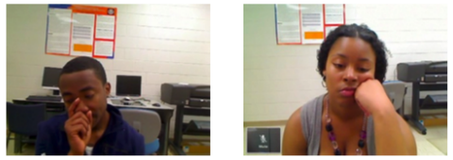
_V_: Cognitive task performance



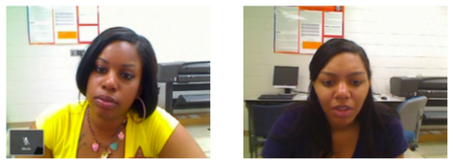
- Whitehill et al. 2014 measured student engagement using (1) observational protocol and (2) automatic engagement detector *d* (q=0.50).
- Using hand-coded labels, corr(*U*, *V*) was estimated as r=0.37.
- Given n, q, r, what is probability of false correlation from d?

Case study: Student engagement vs. cognitive task performance

***U*: Engagement**

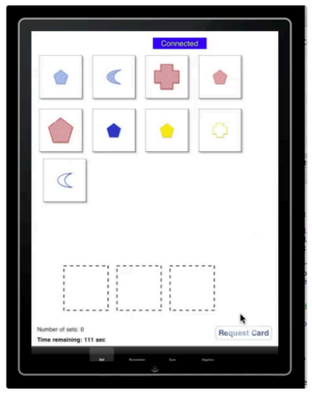


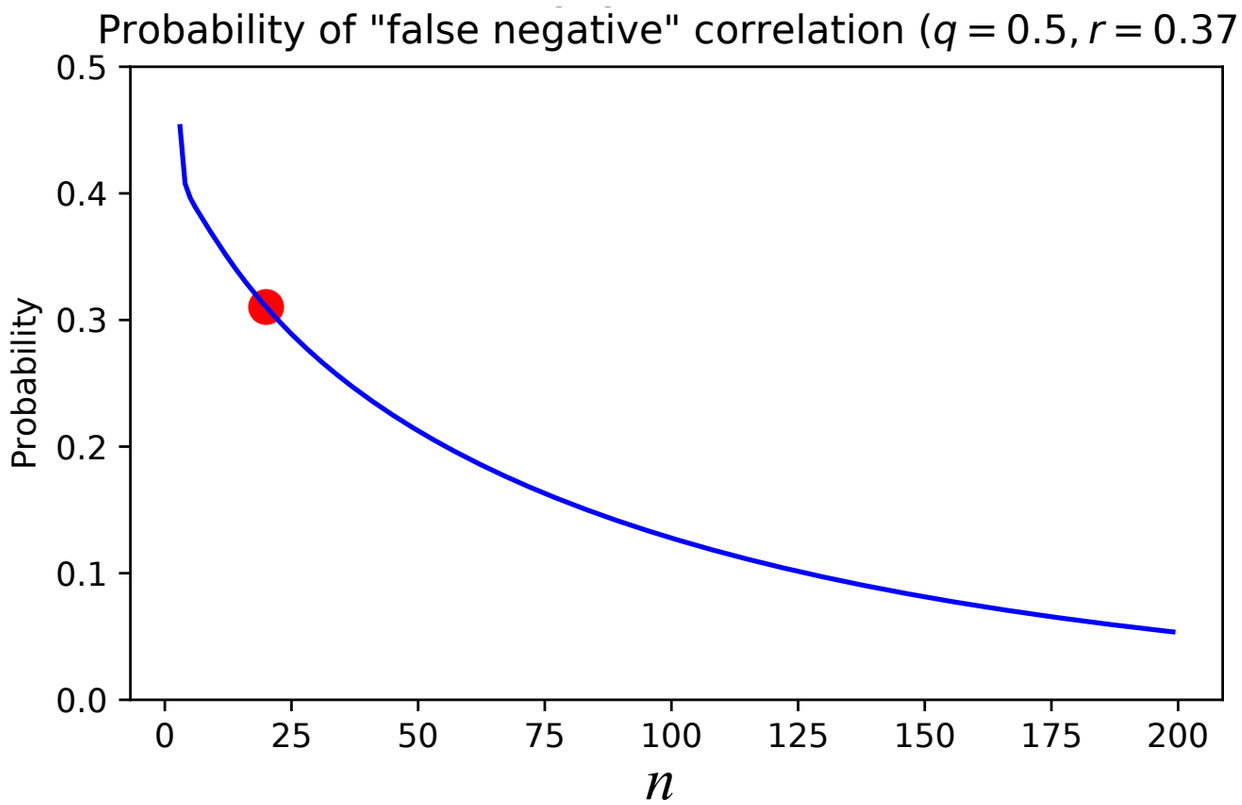
(a) Engagement = 1 (b) Engagement = 2



(c) Engagement = 3 (d) Engagement = 4

***V*: Cognitive task performance**





- Whitehill et al. 2014 measured student engagement using (1) observational protocol and (2) automatic engagement detector *d* (q=0.50).
- Using hand-coded labels, corr(*U*, *V*) was estimated as r=0.37.
- Given n, q, r, what is probability of false correlation from d?