

Using STIX Two in Quarto

Author Name

Sample text

For example, suppose we treat each gene as a *term* and say each sample is a *document*. In that case, we can use word embedding methods in NLP like word2vec or GloVe to learn low-dimensional vector representations for genes, given the gene expression count data for a collection of samples.

$$E \left[(T(F_n) - T(F))^r \right] = E \left[\left(\sum_{j=1}^k T_{j,n} / j! \right)^r \right] + o \left(n^{-(r+k-1)/2} \right)$$

Traditional methods might assume two genes are similar if they have highly correlated expression profiles or overlaps in labels. We assume that two genes are similar if their **expression context** is similar, i.e., the groups of genes that frequently express together with them are similar.

Sample code block

```
#' Classify any file into text or binary
is_text_file <- function(path, n = file.info(path)$size) {
  bytecode <- readBin(path, what = "raw", n = n)

  if (length(bytecode) == 0L) return(FALSE)

  allow <- as.raw(c(9, 10, 13, 32:255))
  block <- as.raw(c(0:6, 14:31))
  cond1 <- any(bytecode %in% allow)
  cond2 <- !any(bytecode %in% block)
  cond1 && cond2
}
```