Worksheet #32: impact of distribution on parallel completion time (rather than locality)

```scala
val points = spark.textFile(…).map(parsePoint).cache()

var w = Vector.random(D) // current separating plane

for (i <- 1 to ITERATIONS) {
  val gradient = points.map(doWork(1)).reduce(_ + _)

  w -= gradient
}

println("Final separating plane: " + w)
```

Consider the above simplified regression program.
Let each doWork operation cost 1 unit of work.
What is the total work? What is the CPL?