Computer Vision Assignment 3

Emilio Brambilla, Lasse Haffke, Moritz Lahann

January 13, 2021

Task 3

The fraction of outliers $\epsilon = \frac{\text{number of outliers}}{\text{total number of data points}}$.

The minimum number of samples needed to define a hypothesis m.

The number of running trials k so that with probability p, at least one set of samples is free from outliers.

ϵ is the probability that a single point is an outlier. Therefore, the probability that a single point is an inlier is $1 - \epsilon$.

If a hypothesis requires multiple samples, the probability of an outlier occurring is multiplied for each sample. For example, if $m = 3$ then the probability of an outlier being among the samples is $p = \epsilon \cdot \epsilon \cdot \epsilon$. Thus, the probability that a single sample of m points contains no outliers is $(1 - \epsilon)^m$.

The inverse of this is the probability that there is an outlier in the sample of m points: $1 - (1 - \epsilon)^m$.

The probability that all k samples drawn fail is $(1 - (1 - \epsilon)^m)^k$.

If p is the probability that at least one set out of k is free from outliers, then the probability that all k samples contain outliers is $1 - p$.

Therefore, $(1 - (1 - \epsilon)^m)^k = (1 - p)$

Since we are interested in k, we can use the logarithm to factor it out.

$$(1-(1-\epsilon)^m)^k = (1-p) \tag{1}$$
$$log(1-(1-\epsilon)^m)^k = log(1-p) \tag{2}$$
$$k \cdot log(1-(1-\epsilon)^m) = log(1-p) \tag{3}$$
$$k = \frac{log(1-p)}{log(1-(1-\epsilon)^m)} \tag{4}$$

We arrive at the conclusion that we can calculate k with the following formula: $k = \frac{log(1-p)}{log(1-(1-\epsilon)^m)}$