Applied Estimation(EL2320) Lab 1 EKF

1   Outliers

Question 7 and the mahalonobis test causes confusion for many students. This is a complicated issue of outlier detection and is related to data association. In practice these are the issues that cause failure.

First off many do not seem to get the relationship between λ_M and δ_M

$$\delta_M = X_2^2(\lambda_M) = \int_0^{\lambda_M} \chi_2^2(u)du \tag{1}$$

In particular since this is a probaility density the integral out to infinity is normalized to 1.0. We see that we can make statments like: 'I want to reject measurements that are in the region so far from expected that the region has a 0.1% probability.'.

That is great but it is hardly the end of this issue since in practice if we expect no outliers and our features are all far from one another compared to our uncertainty then we should just set $\delta_M = 1.0$ and keep all measurements as we would only be throwing away information. As we get more outliers possible we need to do something so that they will not cause filter divergence.

It turns out that seeing an outlier very far from the expected location (relative to our uncertainty as measured by the mahalonobis distance) will have a bad effect if used in the update step. It will both move our mean to a random place and drive our covariance down to an over confident value.

On the other hand if the outlier happens to lie directly where we expected the feature to be its effect will be much less damaging. It would not change our estimated mean at all but would make us overconfident. So in other words the cost of the mistake is not the same for all outliers and that makes this decision complicated.

Added to this is the risk for ambiguous real features so that the MLE data association can give the wrong answer if the feature spacing is small relative to our uncertainty. For that one could say if the two best features have a similar mahalanobis distance we might not want to use the measurement as there is a good chance the data association is wrong.

Question 7 is really not expecting a careful analysis like I am about to do below but rather a disscussion of the effect of the outliers and what can be done about them in general. It might be interesting to see how one might reason for a more modeled answer: Let c denote the situation that the feature is the cause

of the measurement and o denote that it was not the feature (ie an outlier). We do not consider that there could be more than one feature. It might be reasonable to model the situation thus:

$$P(D_M < \lambda)|c) = \frac{X_2^2(\lambda)}{X_2^2(\lambda_{Max})} \tag{2}$$

$$P(D_M < \lambda)|o) = \left(\frac{\lambda}{\lambda_{Max}}\right)^2 \tag{3}$$

$$P(c) = q \tag{4}$$

$$P(o) = (1 - q) \tag{5}$$

Where $X_2^2(\lambda)$ is the cumulative chi square distribution with 2 degrees of freedom. Here we introduce a maximum possible λ value as a sort of size of the 'measurement space'. This is needed as we want to assume the spurious measurements are sort of uniformly spread out in innovation space and thus in the mahalonobis space like the 'area of the circle' (ie quadratically). $1 - q$ is the a priori probability of a spurious measurment.

Bayes then gives us:

$$P(c|D_M < \lambda) = \frac{P(D_M < \lambda_m)|c)q}{P(D_M < \lambda)|c)q + P(D_M < \lambda)|o)(1 - q)} \tag{6}$$

$$P(c|D_M < \lambda) = \left[1 + \frac{1-q}{q}\left(\frac{\lambda}{\lambda_{Max}}\right)^2 \frac{X_2^2(\lambda_{Max})}{X_2^2(\lambda)}\right]^{-1} \tag{7}$$

So we see that if we want to maximize the this there are two competing terms: one the quadratice λ term in the denominator that tends to favor small λ but this is devided by $X_2^2(\lambda)$ which rises rapidly from zero and then flatens out to 1. So for big λ the Quadratic part is changing most and one expects to see the probability to increase as one decreases λ At some point past the 'knee' of X_2^2 one expects this other term to make further decreases in λ cause a decrease the probability.

Thus there should be some λ that is optimal from this point of view. This is found by differentiating:

$$\frac{\partial P(c|D_M < \lambda)}{\partial \lambda} = \frac{1-q}{q}\left(\frac{\lambda}{\lambda_{Max}}\right)^2 \frac{X_2^2(\lambda_{Max})}{X_2^2(\lambda)}(P(c|D_M < \lambda))^2\left[\frac{\chi_2^2(\lambda)}{X_2^2(\lambda)} - \frac{2}{\lambda}\right] = 0 \tag{8}$$

$$\frac{\chi_2^2(\lambda)}{X_2^2(\lambda)} = \frac{2}{\lambda} \tag{9}$$

$$\lambda\chi_2^2(\lambda) = 2\int_0^\lambda \chi_2^2(u)du \tag{10}$$

2

Looking at the plots of the chi square distribution I can see that there is a solution near $\lambda \approx 0.5$. Interesting to note that this did not depend on λ_{Max} or on q. So for this particular model of the outliers and these sensors it would seem 0.5 is about the optimal value for maximimizing the ratio of correct measurements accepted to all accepted measurements.

This ignores another important issue. If we do not accept enough measurements our filter will diverge. So the number of good measurements discarded might be of interest.

It would seem by this analysis that the answer to 7 is more outliers (ie lower q) does not change this threshold. I would say that if one know there are no outliers then one should have a big λ_m to use more measurements. That is you will be throwing away an unreasonble number of them. In some real applications measurements are few and far between and it really hurts to throw away any.

So the real answer is that one must consider the cost of both false positives and false negatives and make some trade off. In practice one adjusts to get the estimator to work on test data.