For both experiments, we choose bandwidth for optimization to be 0.5 and for testing as 0.4. We choose a Gaussian kernel as our weighting function.

We first evaluate the performance of two Gaussian distributions. Samples are drawn from N(0,1) and N(mu,1) and mu varies from 0.15 to 0.75 with step size 0.05. 100 samples are used for optimization and 300 samples for testing. To optimize the choice of theta, we bound theta\_mu and theta\_sigma2 of our weighting function to be [-5,5] and [0.3, +inf). Results of each test are repeated for 500 iterations for alpha = 0.05. In Fig. 1, the power as a function of varying mu is shown. Although they both perform well when the difference is significant, our approach dominates unweighted L2 norm distance when mu is small. In Fig 2. ROC curve for mu=0.3 is shown with varying alpha. Overall, weighted L2 Norm distance performs better than unweighted approach on balancing type I and type II error. For finite samples, even under H0 there can be some minor differences which may be exaggerated through optimization. Fig. 3 shows the empirical Cumulative Distribution Function (CDF) of p-value under H0. Both approaches are close to the uniform distribution which shows that by properly choosing bandwidth and data splitting ratio, our approach can still perform well under H0.

(Fig. 1 power\_normal, Fig. 2 ROC\_normal, Fig. 3 Pvalue)

To emphasize our focus on the locality difference of distribution, we evaluate the performance of Gaussian mixture distribution. Samples are drawn from N(0,1) and 0.9N(0,1)+0.1N(mu,var) and mu varies from 1 to 8 with step size 1. We also use 100 samples for optimization and 300 samples for testing. We bound theta\_mu and theta\_sigma2 to be[-10,10] and [0.3, +inf). Results of each test are repeated for 500 iterations for alpha = 0.05. In Fig. 4(a), unweighted distance has difficulty in distinguishing these two distributions while the weighted distance substantially improves the performance. In Fig 4(b), when the mixture hump is more concentrated, i.e. var=0.1, both approach performs better, but the weighted distance still outperforms unweighted distance. In Fig 5(a-b), we show the ROC curve of two approaches in this experiment. Weighted distance dominantly outperforms the unweighted distance. Intuitively, in Fig. 6(a), we have a mixture hump at mu=7, and after optimization, theta\_mu are very close to the mixture mu and thus amplify the difference nearby. In Fig.6(b), with more concentrated hump, theta\_mu are more close to the mixture mu and the test statistic is much more significant than in Fig. 6(a). Since the unweighted distance treat everywhere equally, it is hard to detect the small difference from a broad scale. The weighted function becomes a magnifier and amplifies the difference between two distributions. Through optimization, locality differences are amplified and thus we can easily detect distant minor changes. Instead of optimization, applicants can actively choose their weighting function parameters to magnify a specific range of interest. In this case, we can imagine that there could possibly be some future applications of this method like change point detection and outlier detection.

(Fig. 4(a) power\_mixturevar11 4(b) power\_mixturevar10\_1, Fig. 5(a) ROC\_mixturevar11 5(b) ROC\_mixturevar10\_1, Fig. 6(a) contour\_mixture\_var11, Fig. 6(b) contour\_mixture\_var10\_1)