Erick Lin & Saurabh Kumar

Homework 3

In this assignment, we attempted to maximize the error of an agnostic mean estimation algorithm over all the points in a sample pertaining to adversarial noise, with the remaining points being drawn from a true distribution. The classes of true distributions we investigated included 1) a spherical Gaussian, 2) a mixture of two spherical Gaussians, 3) the uniform distribution in a simplex, and 4) a mixture of two concentric spherical Gaussians, i.e. two spherical Gaussians with the same mean (covariance matrices set to I with weight 0.2 and 10\*I with weight 0.8. We chose the fourth distribution since we hypothesized that because the two Gaussians had the same mean, putting the noise points at points near the outer Gaussian would shift the mean significantly, but these noise points wouldn’t be ever considered outliers in the outlier removal step since they would be close to points from the outer Gaussian. As a result, we believed we would be able to achieve the highest error rate with this true distribution out of all the distributions. The parameters we varied for all experiments included the number of dimensions and the number of samples. For each situation, we recorded the average error of the algorithm over 10 trials.

For each true distribution, we experimented with four types of noise distributions: 1) the all ones distribution, 2) a constant times the all ones distribution, 3) a constant times the vector (1,0,0,…,0), and 4) a Gaussian distribution with mean (-0.5,0,0,0,…,0) and covariance matrix = 0.5\*I (this one was arbitrary, as we wanted a distribution different than one that was concentrated at one point. For 2) and 3), we optimized for the constant by performing a search between -5 and 5 with a step size of 0.1 and taking the maximum error computed over 10 iterations with each constant value. We chose the range -5 to 5 because if we fixed the distribution at at point whose absolute value was greater than 5, then the error was always low, probably due to the outlier removal step. The results for the different distributions are displayed in the tables below:

**One Spherical Gaussian**

Noise Distribution

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dimensionality | (1,1,1,1,…,1) | c\*(1,1,1,…,1) | c\*(1,0,0,0,…,0) | N(-1, 0.5\*I) |
| 100 | 0.4063 | 0.5732 | 0.6733 | 0.2547 |
| 200 | 0.4366 | 0.5321 | 0.6562 | 0.2401 |
| 500 | 0.3446 | 0.5941 | 0.6742 | 0.2226 |

Values of c that maximized the error for the c\*(1,1,1,…,1) distribution for 100, 200, and 500 dimensions, respectively: -2, -1.4, and -0.5

Values of c that maximized the error for the c\*(1,0,0,…,0) distribution for 100, 200, and 500 dimensions, respectively: -2.7, -2.7, and -2.8

What we found interesting for this distribution was that the value of c that maximized the error using the noise distribution c\*(1,0,0,…,0) was around -2.7 even as we varied the dimension.

**Mixture of Two Gaussians**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dimensionality | (1,1,1,1,…,1) | c\*(1,1,1,…,1) | c\*(1,0,0,0,…,0) | N(-1, 0.5\*I) |
| 100 | 0.7512 | 0.7639 | 0.9206 | 0.5097 |
| 200 | 0.5683 | 0.7115 | 0.9734 | 0.5084 |
| 500 | 0.4279 | 0.7218 | 0.9739 | 0.5030 |

Values of c that maximized the error for the c\*(1,1,1,…,1) distribution for 100, 200, and 500 dimensions, respectively: -1.7, -1.7, and -2

Values of c that maximized the error for the c\*(1,0,0,…,0) distribution for 100, 200, and 500 dimensions, respectively: -2.8, -3, -3

Once again, the value of c that maximized the error using the noise distribution c\*(1,0,0,…,0) was around -3 for different data dimensionality.

**Uniform Distribution in a Simplex**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dimensionality | (1,1,1,1,…,1) | c\*(1,1,1,…,1) | c\*(1,0,0,0,…,0) | N(-1, 0.5\*I) |
| 100 | 0.0021 | 0.0023 | 0.0021 | 0.0016 |
| 200 | 0.0011 | 0.0011 | 0.0011 | 0.0007 |
| 500 | 0.0004 | 0.0004 | 0.0004 | 0.0002 |

Values of c that maximized the error for the c\*(1,1,1,…,1) distribution for 100, 200, and 500 dimensions, respectively: 4.2, 2.4, 5

Values of c that maximized the error for the c\*(1,0,0,…,0) distribution for 100, 200, and 500 dimensions, respectively: 1, 1, 1

**Two Concentric Gaussians**

Note: For the optimization for c in this case, we performed a search from -10 to 10 instead of -5 to 5. This is because the covariance matrix of the outer Gaussian is 10\*I.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dimensionality | (1,1,1,1,…,1) | c\*(1,1,1,…,1) | c\*(1,0,0,0,…,0) | N(-1, 0.5\*I) |
| 100 | 0.7450 | 1.4213 | 1.9846 | 0.6213 |
| 200 | 0.6751 | 1.6070 | 1.9504 | 0.5911 |
| 500 | 0.6669 | 1.5028 | 1.9732 | 0.5977 |

Values of c that maximized the error for the c\*(1,1,1,…,1) distribution for 100, 200, and 500 dimensions, respectively: 8, -0.5, 9.5

Values of c that maximized the error for the c\*(1,0,0,…,0) distribution for 100, 200, and 500 dimensions, respectively: 8.2, -8, -8.1

**Overall Analysis:**

According to our intuition, the error is maximized only when the noisy points are never removed in any of the outlier removal steps, and furthermore, all share the same position. This is because if their positions are different, then they are more likely to be removed as outliers, and otherwise, they collectively have a smaller influence on the sample mean because they destructively interfere with one another. With this assumption, the location in which we decided to place the noisy points depended on which true distribution was being used. Due to the symmetry in distributions 1 and 4, only the distance, not the direction, of this location from the mean of the points sampled from the true distribution influences the error value. However, for distribution 2, the error is maximized when the noisy points lie in the direction of the mean of the Gaussian of smaller weight relative to the mean of the Gaussian of larger weight, since in this case the effects of the noisy points and the points sampled from the Gaussian of smaller weight on the error compound one another. For distribution 3, using the fact that the true mean coincides with the centroid of the simplex, the error is maximized when the noisy points lie in the direction of the vertex that is farthest from the centroid, since this case minimizes the number of points sampled from the true distribution that cancel out the effect of the noisy points on the error. For distribution 4, the error is maximized when the noise points are placed at near where the outer Gaussian distribution’s points are concentrated.