

---

## Discussion 3

---

**Note:** Your TA will probably not cover all the problems on this worksheet. The discussion worksheets are not designed to be finished within an hour. They are deliberately made slightly longer so they can serve as resources you can use to practice, reinforce, and build upon concepts discussed in lectures, discussions, and homework.

**This Week's Cool AI Demo/Video:**

<https://youaretheassistantnow.com/>

<https://gemini.google.com/>

### 1 MLE with a Linear Constraint on Independent Gaussians

(a) Let  $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$  and  $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$  be independent. You only observe their sum

$$S := X + Y.$$

Find the Maximum Likelihood Estimate (MLE) of  $X$  given  $S = 2$ .

(i) Show that, up to a proportionality constant,

$$p_{X|S}(x | 2) \propto p_X(x) p_Y(2 - x).$$

(ii) Use the Gaussian forms to show that maximizing  $p_{X|S}(x | 2)$  with respect to  $x$  is equivalent to minimizing

$$\frac{(x - \mu_X)^2}{2\sigma_X^2} + \frac{(2 - x - \mu_Y)^2}{2\sigma_Y^2}.$$

(iii) Find the estimate,  $\hat{x}_{\text{MLE}}$ , that maximizes the likelihood  $p_{X|S}(x | 2)$  using the expression you found in (ii). Simplify to a weighted-average form.

(iv) Consider the special case where  $\mu_X = \mu_Y = 0$  and  $\sigma_X^2 = \sigma_Y^2$ . How can we simplify the maximum likelihood estimate?

## 2 Binomial MLE with Misclassification

(a) You run  $m$  independent experiments. In each experiment, you perform  $n$  independent Bernoulli trials with unknown true success probability  $p$ . However, your measurement device is imperfect:

- A true success is recorded as a success with probability  $1 - q$  (and as a failure with probability  $q$ ).
- A true failure is recorded as a success with probability  $r$  (and as a failure with probability  $1 - r$ ).

Let  $X_i \sim \text{Bin}(n, p)$  denote the unobserved number of true successes in experiment  $i$ , and let  $Y_i$  be the observed number of recorded successes. Recall that the probability mass function for the binomial random variable  $X_i \sim \text{Bin}(n, p)$  is

$$p(x_i) = \binom{n}{x_i} p^{x_i} (1 - p)^{n - x_i}.$$

Derive the likelihood for the observations  $\mathcal{D} = \{y_i\}_{i=1}^m$ . *Hint: Your answer should be a function of the true success probability  $p$ .*

(b) Find the MLE  $\hat{p}$  in closed form. Show all steps.

### 3 Proof of K-means Convergence

Consider the K-means algorithm applied to data points  $\{x_n\}_{n=1}^N \subset \mathbb{R}^D$  with  $K$  clusters. Define binary indicator variables  $r_{nk} \in \{0, 1\}$  with  $\sum_{k=1}^K r_{nk} = 1$  for each  $n$ , and cluster centers  $\{\mu_k\}_{k=1}^K$ . The K-means objective is

$$J(\{r_{nk}\}, \{\mu_k\}) = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2.$$

The K-means algorithm alternates between

1. **Update assignments:** Set  $r_{nk} = 1$  for the  $k$  that minimizes  $\|x_n - \mu_k\|^2$  and  $r_{nj} = 0$  for all  $j \neq k$ .
2. **Update means:** Set the cluster centers to

$$\mu_k = \frac{\sum_{n=1}^N r_{nk} x_n}{\sum_{n=1}^N r_{nk}} \quad (\text{for clusters with } \sum_n r_{nk} > 0).$$

Prove that the K-means algorithm converges after a finite number of iterations. (Hint: It is sufficient to show that the objective is monotonically decreasing, and that there are a finite number of iterations).