Exercise 1: LogSumExp Trick

a) Numerical Instability in Computing p^* or p^+

We define:

$$p^* = \prod_{i=1}^{n} p_i, \qquad p^+ = \sum_{i=1}^{n} p_i$$

where $p_i \in [0, 1]$ are probabilities.

Computing p^* directly involves multiplying many small probabilities:

$$p^* = \prod_{i=1}^{n} p_i$$

If $p_i = 10^{-10}$ and $n = 100$, then:

$$p^* = (10^{-10})^{100} = 10^{-1000} \approx 0$$

This underflows to zero in floating-point arithmetic.

For p^+:

$$p^+ = \sum_{i=1}^{n} p_i$$

If $p_1 = 10^{-10}$ and $p_2 = 1.0$, then:

$$p^+ \approx 1.0$$

The small term disappears due to precision loss.

Final Thought: Direct computation can lead to underflow or loss of precision.

b) Compute $\ln p^*$ in log-space

$$\ln p^* = \ln \left(\prod_{i=1}^{n} p_i \right) = \sum_{i=1}^{n} \ln p_i$$

Thus:

$$\boxed{\ln p^* = \sum_{i=1}^{n} \ln p_i}$$

c) Compute $\ln p^+$ using LogSumExp trick

Original expression:

$$\ln p^+ = \ln \left(\sum_{i=1}^{n} \exp(\ln p_i) \right)$$

i) Apply LogSumExp trick: Let:

$$c = \max_i \ln p_i$$

Then:

$$\ln p^+ = \ln\left(\sum_{i=1}^{n} \exp(\ln p_i - c)\cdot \exp(c)\right) \tag{1}$$
$$= \ln\left(\exp(c)\sum_{i=1}^{n} \exp(\ln p_i - c)\right) \tag{2}$$
$$= c + \ln\left(\sum_{i=1}^{n} \exp(\ln p_i - c)\right) \tag{3}$$

So:

$$\boxed{\ln p^+ = c + \ln\left(\sum_{i=1}^{n} \exp(\ln p_i - c)\right)}$$

ii) Why does this help? Subtracting c ensures all exponents are ≤ 0, so $\exp(\ln p_i - c) \in [0, 1]$. This prevents overflow when $\ln p_i$ is large and underflow when $\ln p_i$ is very negative.

Exercise 2: Gaussian Kernel is a Mercer Kernel

The Gaussian kernel is defined as:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

Proof:

1. Expand $\|x - y\|^2$:

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2x^\top y$$

So:

$$K(x, y) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right)\exp\left(-\frac{\|y\|^2}{2\sigma^2}\right)\exp\left(\frac{x^\top y}{\sigma^2}\right)$$

2. Expand $\exp(x^\top y/\sigma^2)$ using Taylor series:

$$\exp(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!}$$

Thus:

$$\exp\left(\frac{x^\top y}{\sigma^2}\right) = \sum_{k=0}^{\infty} \frac{(x^\top y)^k}{k!\sigma^{2k}}$$

3. Apply multinomial theorem:

$$(x^\top y)^k = \left(\sum_{i=1}^{d} x_i y_i\right)^k = \sum_{\alpha_1+\cdots+\alpha_d=k} \frac{k!}{\alpha_1!\ldots\alpha_d!} \prod_{i=1}^{d} (x_i^{\alpha_i} y_i^{\alpha_i})$$

4. So:

$$K(x,y) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|y\|^2}{2\sigma^2}\right) \sum_{k=0}^{\infty} \sum_{\alpha} \frac{1}{\alpha_1!\ldots\alpha_d!\sigma^{2k}} \prod_{i=1}^{d} x_i^{\alpha_i} y_i^{\alpha_i}$$

5. This is an inner product:

$$K(x,y) = \langle \phi(x), \phi(y) \rangle$$

where $\phi(x)$ is an infinite-dimensional feature map:

$$\phi(x) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) \left(\sqrt{\frac{1}{\alpha_1!\ldots\alpha_d!\sigma^{2k}}} \prod_{i=1}^{d} x_i^{\alpha_i}\right)_{\forall \alpha}$$

Final Thought: Since $K(x,y)$ can be expressed as an inner product in a Hilbert space, the Gaussian kernel is a Mercer kernel.