AML-Quiz-10

March 24, 2022

Question 1

Consider the Low-rank Gaussian Process Parameterization method for multivariate forecasting.

Let there be two time series $\mathbf{z_1}$ and $\mathbf{z_2}$ where z_{it} denotes the value of series i at time t.

Part a: [1 mark] [The first step in the Copula paper is to transform the values in each series to a standard Gaussian. Suppose for series 1 we observe $\mathbf{z_1} = \{0, 3, 2, 1, 5.5, 5\}$, which of the following shows the corresponding $\mathbf{x_1}$ series. Use $\Phi^{-1}(f)$ to denote the inverse of the CDF function of the standard Gaussian.

1. $\Phi^{-1}(1/6), \Phi^{-1}(4/6), \Phi^{-1}(3/6), \Phi^{-1}(2/6), \Phi^{-1}(1)\Phi^{-1}(5/6)$
2. $\Phi^{-1}(0), \Phi^{-1}(3/6), \Phi^{-1}(2/6), \Phi^{-1}(1/6), \Phi^{-1}(5/6)\Phi^{-1}(4/6)$
3. $\Phi^{-1}(0), \Phi^{-1}(6/16.5), \Phi^{-1}(3/16.6), \Phi^{-1}(1/16.5)\Phi^{-1}(1), \Phi^{-1}(11.5/16.5)$
4. $\Phi^{-1}(1/6), \Phi^{-1}(4/6), \Phi^{-1}(3/6), \Phi^{-1}(2/6)\Phi^{-1}(5/6), \Phi^{-1}(1)$

Answer: The GP paper first creates an empirical CDF function as

$$\hat{F}_i(x) = \frac{1}{m} 1_{z_{it} \leq x}$$

Using that we obtain the CDF for the 5 values as [1/6, 4/6, 3/6, 2/6, 6/6,5/6]. The corresponding x values are then $\Phi^{-1}(.)$ of the above values

Part b: [1 mark] Let $\hat{x_{t7}} = 0$ denote the predicted mean value. Convert this into the predicted $\hat{z_{t7}}$ value.

Answer: We know that for standard Gaussian at 0, the CDF is 0.5. The z at this CDF from the above calculation is at 2. So the actual predicted value is 2.

Part c: [1 mark] Next we attempt to model the $\mathbf{x_{\cdot t}}$ joint distribution using a low-rank Gaussian. Let h_{it} denote the RNN state for the i^{th} time series at time

t. The vector $\mathbf{y_{it}}$ is defined as $\mathbf{y_{it}} = [\mathbf{h_{it}}, \mathbf{e_i}]$, where $\mathbf{e_i}$ is some series dependant feature. It is known that at time t, $\mathbf{y_{1t}} = [0, -1, 1]; \mathbf{y_{2t}} = [2, -2, 0]$.

2 linear neural networks d_θ and v_ϕ are used to generate the diagonal and the low rank matrix respectively.

Consider $d_\theta(\mathbf{y_{it}}) = y_{it0} - y_{it1} + y_{it2}$, and $v_\phi(\mathbf{y_{it}}) = y_{it0} - y_{it1} + 2y_{it2}$. If $\Sigma(\mathbf{y_t})$ denotes the kernel matrix at time t, then $\det \mathbf{\Sigma}(\mathbf{y_t})$ is:

Answer: 76
$d_1 = 0+1+1 = 2; d_2 = 2+2+0 = 4$. Also $v_1 = 0+1+2 = 3; v_2 = 2+2+0 = 4$.
So $D + VV^T = [[11, 12], [12, 20]]$

Part d: [1 mark - MCQ] For this part, assume that we have access to only the first time series, i.e. we only know $\mathbf{x_{1t}}$ and h_{1t}. Using the same neural networks as in Part c, if $x_t \sim GP(\mathbf{0}, \Sigma(\mathbf{y_t}))$, and $x_{1t} = 0.5$ then $\mu(x_{2t})$ is:

Answer: 6/11
$d_1 = 0 + 1 + 1 = 2$ and $v_1 = 0 + 1 + 2 = 3$. So $\sigma^2(y_1) = 2 + 9 = 11$
$v_2 = 2 + 2 + 0 = 4$
Mean $= \mathbf{k}_*^T \Sigma^{-1} z$. $\mathbf{k}_* = (3 \times 4/11) * x_{1t} = 6/11$