Gaussian Process (GP) on Composite SEDs

For a given type of composite SED,

$$\underline{f}_\lambda \sim \mathcal{GP}(\underline{\mu}(\underline{f}^{(\text{rest})}, z), \underline{\underline{K}}), \tag{1}$$

where

$$\underline{f}^{(\text{obs})}(\underline{\lambda}) = \underline{f}^{(\text{rest})}(\frac{\lambda}{1+z}).$$

GP is fully specified by two quantities, the mean function $\underline{\mu}$ and the covariance $\underline{\underline{K}}$.

$$\begin{aligned} \mu(\lambda) &= \mathbb{E}[f^{(\text{rest})}(\lambda) \mid \lambda] \\ \underline{\underline{K}} &= \text{cov}[f^{(\text{rest})}(\lambda), f^{(\text{rest})}(\lambda') \mid \lambda, \lambda'] \end{aligned} \tag{2}$$

Given a finite set of $\underline{\lambda}$, the $\underline{f}$ would be multivariate Gaussian distributed,

$$\begin{aligned} p(\underline{f}) &= \mathcal{N}(\underline{f}; \mu(\underline{\lambda}), K(\underline{\lambda}, \underline{\lambda})) \\ &= \frac{1}{\sqrt{(2\pi)^d \text{det}\underline{\underline{K}}}} \exp\left(-\frac{1}{2}(\underline{f} - \underline{\mu})^{\mathrm{T}} \underline{\underline{K}}^{-1} (\underline{f} - \underline{\mu}).\right) \end{aligned}$$

The choice of $\underline{\underline{K}}$ reflects the covariance structure of the composite SED with a given galaxy type. The prior for noise could be brought into the GP using variances $\underline{v} = \sigma(\underline{\lambda})^2$ of the measurements from the flux,

$$p(\underline{f}) = \mathcal{N}(\underline{f}; \mu(\underline{\lambda}), K(\underline{\lambda}, \underline{\lambda}), \underline{\underline{V}})$$

where $\underline{\underline{V}} = \text{diag}\underline{v}$.

If necessary, we can model an extra noise term $\underline{\underline{\Omega}} = \text{diag}\underline{\omega}$ which depends on the redshift.

The complete prior expression,

$$p(\underline{y} \mid \underline{f}, \underline{v}, z, \mathcal{M}_i) = \mathcal{N}(\underline{f}; \mu(\underline{\lambda}), K(\underline{\lambda}, \underline{\lambda}), \underline{\underline{V}}, \underline{\underline{\Omega}}) \tag{3}$$

where $\mathcal{M}_i$, $i \in \{\text{QG}, \text{PSB}, \text{TG}, \text{SFG}, \text{(D)SFG}, \text{ELG}\}$, corresponds to different galaxy types identified by composite SEDs and human justifications.

To learn the mean vector $\underline{\mu}$, simply just take the mean at rest frame,

$$\underline{\mu}_i = \text{median}_{\neg\text{NaN}}(\underline{y}_i). \tag{4}$$

For the positive semi-definite covariance matrix $\underline{\underline{K}}$, decompose it into $\underline{\underline{K}} = \underline{\underline{M}}\,\underline{\underline{M}}^{\mathrm{T}}$, and train the $\underline{\underline{M}}$ based on taking first few eigenspectra of principal component analysis (PCA).

If we have any additional assumption on the redshift dependent noise terms need to be specified, we can use maximum likelihood estimation to get the optimal parameters for noise.

Bayesian Model Selection

Bayesian model selection formula,

$$\Pr(\mathcal{M} \mid \mathcal{D}) = \frac{\Pr(\mathcal{D} \mid \mathcal{M})\Pr(\mathcal{M})}{\sum_i \Pr(\mathcal{D} \mid \mathcal{M}_i)\Pr(\mathcal{M}_i)}.$$

Let i be some pre-selected galaxy types, i.e., $i = ...$

1. QG : quiescent galaxy
2. PSB : post-starburst galaxy
3. TG : transitioning galaxy
4. SFG : star-forming galaxy
5. (D)SFG : (dusty) star-forming galaxy
6. ELG : emission-line galaxy

Usually, the prior depends on the redshift, so

$$\Pr(\mathcal{M} \mid \mathcal{D}, z) = \frac{\Pr(\mathcal{D} \mid \mathcal{M}, z)\Pr(\mathcal{M} \mid z)}{\sum_i \Pr(\mathcal{D} \mid \mathcal{M}_i, z)\Pr(\mathcal{M}_i \mid z)}, \tag{5}$$

where $i \in \{\mathrm{QG}, \mathrm{PSB}, \mathrm{TG}, \mathrm{SFG}, \mathrm{(D)SFG}, \mathrm{ELG}\}$.

$\Pr(\mathcal{M} \mid \mathcal{D}, z)$ would give a probabilistic prediction of the classification of the incoming photometric data.

The marginal cases would give roughly even probabilities on each $\mathcal{M}_i$. We could just keep them in another place and analyze them later in the future. Usually, we expect to find new classification types in the marginal cases.

Issue: since our GP is trained on composite SEDs with discrete pixels, how to deal with missing pixels (sparse photometric spectra) would be a problem.