Lecture 6: Gaussian process for learning

Tianpei Xie

Jul. 15th., 2015

Contents

1 Definitions **2**
1.1 Gaussian process with feature space as index 2
1.2 Covariance function 3

1 Definitions

1.1 Gaussian process with feature space as index

- Let $f : \mathcal{X} \to \mathcal{Y}$ be a measureable functions in a RKHS $\mathcal{H} \subset \mathcal{Y}^{\mathcal{X}}$ associated with kernel K, where $\mathcal{X} \subset \mathbb{R}^d$ is the feature domain and $\mathcal{Y} \subset \mathbb{R}$ is the decision space. Define a linear functional $\eta : \mathcal{H} \to \mathbb{R},\ f \mapsto \eta(f)$ and $\eta \in \mathcal{H}^* \simeq \mathcal{H}$, the space of all linear functionals. In specific, $\{\phi_i\}$ is the set of eigenfunctions of K associated with eigenvalue $\{\lambda_i\}$, which also forms a set of orthonormal basis in $\mathcal{H}$,

$$
\begin{aligned}
\lambda_i \phi_i(x) &= \langle K(\cdot, x)\,,\, \phi_i \rangle = \int_{\mathcal{X}} \phi_i(z) K(z, x) d\mu(z), \forall x \in \mathcal{X} \\
K(x, x') &= \sum_i \lambda_i \phi_i(x) \phi_i(x') \\
f &= \sum_i \beta_i \phi_i = \sum_i e_i \sqrt{\lambda_i} \phi_i, \qquad \sum_{i=1}^{\infty} \beta_i^2 / \lambda_i = \sum_{i=1}^{\infty} e_i^2 < \infty \\
&= \sum_m \widehat{\beta}_m K(\cdot, x_m) \\
\eta(\cdot) &= \langle \cdot\,,\, \eta \rangle = \sum_i \alpha_i \langle \cdot\,,\, \phi_i \rangle_{\mathcal{H}} = \sum_n \widehat{\alpha}_n \langle \cdot\,,\, K(\cdot, x_n) \rangle \\
\eta(f) &= \sum_i \alpha_i \beta_i = \sum_{n,m} \widehat{\alpha}_n \widehat{\beta}_m K(x_n, x_m) \\
\langle f\,,\, g \rangle_{\mathcal{H}} &= \sum_{i=1}^{\infty} \frac{f_i g_i}{\lambda_i} = \left\langle K^{-1} f\,,\, g \right\rangle, \\
f(x) &= \langle f\,,\, K(\cdot, x) \rangle_{\mathcal{H}}\,.
\end{aligned}
$$

- A random function on feature domain is given by $f : \mathcal{X} \times \Omega \to \mathbb{R}$. Note that the index set is the feature domain $\mathcal{X}$ not the conventional time domain T. Assume that the domain space $\mathcal{X}$ is Hausdorff, locally convex and separable so that the results in previous sections hold in general.

- A random function f can be seen as generated by the while noise Gaussian measure (Wiener measure) $\mathcal{W}$ on $\ell^2 \subset \mathbb{R}^\infty$.

  Let $e \equiv (e_i, i = 1, \ldots) \in \ell^2$ with $\sum_i^\infty e_i^2 < \infty$. Then a white noise Gaussian measure $\mathcal{W}(e)$ has zero mean and

$$
\int_{\ell^2} e_i e_j \mathcal{W}(de) = \delta_{i,j} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}
$$

  Then $f \sim \mathcal{G}(\mathcal{H})$, if and only if

$$
f(\cdot) = \sum_i \sqrt{\lambda_i} \phi_i(\cdot) e_i
$$

where $\{\phi_i\}$ is the set of eigenfunctions of K associated with eigenvalue $\{\lambda_i\}$, with respect to Lebesgue measure μ on $\mathcal{X}$.

- The function space $\mathcal{H}$ is equipped with σ-algebra $\mathscr{B}$ generated by the collection of all cylinder sets $\{f \in \mathcal{H} : (\eta_1(f), \ldots, \eta_k(f)) \in A\}$ for all k, all $\eta_1, \ldots, \eta_k \in \mathcal{H}^*$ are linear functionals on $\mathcal{H}$ and all $A \in \mathcal{B}(\mathbb{R}^k)$. A induced probability measure $\mathcal{P} \equiv \mathbb{P} \circ f_{\cdot}^{-1}$ defined on $\mathscr{B}$ is given as

$$\mathcal{P}(C) \equiv \mathbb{P}\left\{\omega : f_{\cdot} \equiv f(\cdot, \omega) \in C\right\},\ C \in \mathscr{B}$$

- In practice, one could define a sampling map $\mathcal{S} : T \to \mathcal{X}$ that induced a sampling ordering from T over the field $\mathcal{X}$, then the sample path is $f(\boldsymbol{x}_t, \omega)$ not $f(t, \omega)$ for $t \in T$. Since $\mathcal{X}$ is separable, the image $\overline{\mathcal{S}(T)} = \mathcal{X}$ is dense.

  We may consider a random function $g : T \times \Omega \to \mathbb{R}$ with sample path $g(\cdot, \omega) = f(\cdot, \omega) \circ \mathcal{S}$ as a conventional process.

- Given $\mathcal{H}$, the random function $f \sim \mathcal{G}(\mathcal{H})$, the Gaussian measure on function space $\mathcal{H}$, if and only if all its linear functionals $\eta(f) \in \mathcal{H}^*$ yields a Gaussian distribution on $\mathbb{R}$.

1.2 Covariance function

- The dual space $\mathcal{H}^*$ has all linear functional $I(f)$. Note that the evaluation functional of f at ξ is a linear functional, as $\xi(f) \equiv f(\xi)$.

- The linear operator $K : \mathcal{H} \to \mathcal{H}$ is called the *covariance operator* of a measure $\mathcal{P}$ if for any $\xi, \eta \in \mathcal{H}^* \simeq \mathcal{H}$, the following equality holds,

$$\xi\left(K(\eta)\right) = \int_{\mathcal{H}} \xi(f - m)\eta(f - m)\mathcal{P}(df)$$

- Note that $\xi(f) \equiv f(\xi) = \langle f\, ,\, K(\cdot, \xi)\rangle_{\mathcal{H}} \in \mathcal{H}^*$ and $\eta(f) \equiv f(\eta) = \langle f\, ,\, K(\cdot, \eta)\rangle_{\mathcal{H}} \in \mathcal{H}^*$ are two functionals on $\mathcal{H}$. Therefore,

$$
\begin{aligned}
cov(f(\xi), f(\eta)) \equiv \xi\left(K(\eta)\right) &= \int_{\mathcal{H}} \xi(f)\eta(f)\mathcal{P}(df) \\
&= \int_{\mathcal{H}} \langle f\,,\, K(\cdot,\xi)\rangle_{\mathcal{H}} \langle f\,,\, K(\cdot,\eta)\rangle_{\mathcal{H}} \mathcal{P}(df) \\
&= \int_{\mathbb{R}^\infty} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \beta_i\beta_j \langle \phi_i\,,\, K(\cdot,\xi)\rangle_{\mathcal{H}} \langle \phi_j\,,\, K(\cdot,\eta)\rangle_{\mathcal{H}} \mathcal{W}(d\boldsymbol{\beta}) \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \left(\int_{\mathbb{R}^\infty} \beta_i\beta_j \mathcal{W}(d\boldsymbol{\beta})\right) \phi_i(\xi)\phi_j(\eta) \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \lambda_i\delta_i(j)\, \phi_i(\xi)\phi_j(\eta) \\
&= \sum_{i=1}^{\infty} \lambda_i\phi_i(\xi)\phi_i(\eta) \\
&= K(\xi,\eta)
\end{aligned}
$$

where $\mathcal{W}(d\boldsymbol{\beta}) = \mathcal{N}(0, \mathrm{diag}(\lambda_1, \ldots,))d\boldsymbol{\beta}$ so that $\sum_{i=1}^{\infty} \beta_i^2/\lambda_i < \infty$