A Kernel for Hierarchical Parameter Spaces

August 23, 2013

Consider a D-dimensional input space $\mathcal{X}$ with an associated DAG structure $\mathcal{D}$ over the dimensions $1, \ldots, D$, where each input dimension is only active under some instantiations of its ancestor dimensions in $\mathcal{D}$. We specify a kernel for $\mathcal{X}$ as a product over individual kernels for each dimension i. Each of these individual kernels is specified using an isometric embedding into a Euclidean space ($\mathbb{R}^2$ for continuous-valued parameters, and $\mathbb{R}^m$ for categorical-valued parameters with m choices). We restrict dimensions that are parents in $\mathcal{D}$ to be categorical-valued. Leaf dimensions in $\mathcal{D}$ can be continuous-valued or categorical-valued.

1 Continuous Dimensions

Let's first consider a continuous-valued input dimension i with upper and lower bounds U_i and L_i, respectively. We define a delta function δ_i that maps complete inputs $\underline{x} \in \mathcal{X}$ to true if dimension i is active in the context of the instantiation of i's ancestors in $\underline{x}$, and to false otherwise. Let d_E denote the Euclidean distance metric in $\mathbb{R}^2$. We define a metric ${d_i}^{(cont)}$ on $\mathcal{X}$ and an isometric embedding ${f_i}^{(cont)}$ of $(\mathcal{X}, {d_i}^{(cont)})$ into $(\mathbb{R}^2, d_E)$:

$$
{d_i}^{(cont)}(\underline{x}, \underline{x}') = \left\{ \begin{array}{ll} 0 & \text{if } \delta_i(\underline{x}) = \delta_i(\underline{x}') = \text{false} \\ w_i & \text{if } \delta_i(\underline{x}) \neq \delta_i(\underline{x}') \\ w_i\sqrt{2}\sqrt{1 - \cos(\pi \frac{x_i - x_i'}{U_i - L_i})} & \text{otherwise.} \end{array} \right\}
$$

$$
{f_i}^{(cont)}(\underline{x}) = \left\{ \begin{array}{ll} [0, 0]^\intercal & \text{if } \delta_i(\underline{x}) = \text{ false} \\ w_i[\sin \pi \frac{x_i}{U_i - L_i}, \cos \pi \frac{x_i}{U_i - L_i}]^\intercal & \text{otherwise.} \end{array} \right\}
$$

Proposition 1. *Embedding ${f_i}^{(cont)}$ is an isometric embedding of $(\mathcal{X}, {d_i}^{(cont)})$ into $(\mathbb{R}^2, d_E)$.*

Proof. Consider two inputs $\underline{x}, \underline{x}' \in \mathcal{X}$. We need to show that ${d_i}^{(cont)}(\underline{x}, \underline{x}') = d_E({f_i}^{(cont)}(\underline{x}), {f_i}^{(cont)}(\underline{x}'))$. We use the abbreviation $\alpha = \pi \frac{x_i}{U_i - L_i}$ and $\alpha' = \pi \frac{x_i'}{U_i - L_i}$ and consider the following three possible cases of dimension i being active or inactive in $\underline{x}$ and $\underline{x}'$.

Case 1: $\delta_i(\underline{x}) = \delta_i(\underline{x}') =$ false. In this case, we trivially have

$$d_E({f_i}^{(cont)}(\underline{x}), {f_i}^{(cont)}(\underline{x}')) = d_E([0,0]^\intercal, [0,0]^\intercal) = 0 = {d_i}^{(cont)}(\underline{x}, \underline{x}').$$

Case 2: $\delta_i(\underline{x}) \neq \delta_i(\underline{x}')$. In this case, we have

$$d_E({f_i}^{(cont)}(\underline{x}), {f_i}^{(cont)}(\underline{x}')) = d_E([\sin\alpha, \cos\alpha]^\intercal, [0,0]^\intercal) = \sqrt{w_i^2(\sin^2\alpha + \cos^2\alpha)} = w_i = d_i(\underline{x}, \underline{x}'),$$

and symmetrically for $d_E([0,0]^\intercal, [\sin\alpha, \cos\alpha]^\intercal)$.

Case 3: $\delta_i(\underline{x}) = \delta_i(\underline{x}') =$ true. We have:

$$\begin{aligned}
d_E({f_i}^{(cont)}(\underline{x}), {f_i}^{(cont)}(\underline{x}')) &= d_E(w_i[\sin\alpha, \cos\alpha]^\intercal, w_i[\sin\alpha', \cos\alpha']^\intercal) \\
&= w_i\sqrt{(\sin\alpha - \sin\alpha')^2(\cos\alpha - \cos\alpha')^2} \\
&= w_i\sqrt{\sin^2\alpha - 2\sin\alpha\sin\alpha' + \sin^2\alpha' + \cos^2\alpha - 2\cos\alpha\cos\alpha' + \cos^2\alpha'} \\
&= w_i\sqrt{(\sin^2\alpha + \cos^2\alpha) + (\sin^2\alpha' + \cos^2\alpha') - 2(\sin\alpha\sin\alpha' + \cos\alpha\cos\alpha')} \\
&= w_i\sqrt{1 + 1 - 2\cos(\alpha - \alpha')} \qquad (1) \\
&= w_i\sqrt{2}\sqrt{1 - \cos(\pi\frac{x_i - x_i'}{U_i - L_i})} = d_i(\underline{x}, \underline{x}'),
\end{aligned}$$

where (1) follows from the previous line by using the identity

$$\cos(a - b) = \cos a \cos b + \sin a \sin b.$$

$\square$

Define $\kappa\colon \mathbb{R}^+ \to \mathbb{R}$ as a semi-positive definite function over Euclidean space. By this, we mean that $K \in \mathbb{R}^{N\times N}$, defined by

$$K_{m,n} = \kappa\big(d_E(\underline{y}_m, \underline{y}_n)\big), \quad \text{for } \underline{y}_m, \underline{y}_n \in \mathbb{R}^P, \quad m, n = 1, \ldots, N, \qquad (2)$$

is positive semi-definite. A popular example of such a κ is the exponentiated quadratic, for which $\kappa(\delta) = \sigma^2\exp(-\frac{1}{2}\frac{\delta^2}{\lambda^2})$; another popular choice is the rational quadratic, for which $\kappa(\delta) = \sigma^2(1 + \frac{1}{2\alpha}\frac{\delta^2}{\lambda^2})^{-\alpha}$.

Proposition 2. $\kappa\big({d_i}^{(cont)}(\cdot,\cdot)\big)$ *is a positive semi-definite covariance function over input space* $\mathcal{X}$.

Proof. We require that $K \in \mathbb{R}^{N\times N}$ defined by

$$K_{m,n} = \kappa\big({d_i}^{(cont)}(\underline{x}_m, \underline{x}_n)\big), \quad \text{for } \underline{x}_m, \underline{x}_n \in \mathcal{X}, \quad m, n = 1, \ldots, N,$$

is positive semi-definite. Now, from above,

$$\begin{aligned}K_{m,n} &= \kappa\Big(d_E({f_i}^{(cont)}(\underline{x}_m), {f_i}^{(cont)}(\underline{x}_n))\Big) \\ &= \kappa\big(d_E(\underline{y}_m, \underline{y}_n)\big)\end{aligned}$$

where $\underline{y}_m = {f_i}^{(cont)}(\underline{x}_m)$ and similar for $\underline{x}_m$. Then, by assumption that κ is semi-positive definite function over Euclidean space, K is positive semi-definite. $\square$

2 Categorical Dimensions

Next, let's consider a categorical-valued input dimension i with m_i possible values $V_i = \{v_{i,1}, \ldots, v_{i,m_i}\}$. As in the continuous case, we define a delta function δ_i that maps complete inputs $\underline{x} \in \mathcal{X}$ to true if dimension i is active in the context of the instantiation of i's ancestors in $\underline{x}$, and to false otherwise. Let ${d_E}^{m_i}$ denote the Euclidean distance metric in $\mathbb{R}^{m_i}$. We define a metric ${d_i}^{(cat)}$ on $\mathcal{X}$ and an isometric embedding of $(\mathcal{X}, {d_i}^{(cat)})$ into $(\mathbb{R}^{m_i}, {d_E}^{m_i})$:

$${d_i}^{(cat)}(\underline{x}, \underline{x}') = \left\{ \begin{array}{ll} 0 & \text{if } \delta_i(\underline{x}) = \delta_i(\underline{x}') = \text{false} \\ w_i & \text{if } \delta_i(\underline{x}) \neq \delta_i(\underline{x}') \\ w_i\sqrt{2}\mathbb{I}_{x_i \neq x_i'} & \text{otherwise.} \end{array} \right\}$$

$${f_i}^{(cat)}(\underline{x}) = \left\{ \begin{array}{ll} \underline{0} \in \mathbb{R}^{m_i} & \text{if } \delta_i(\underline{x}) = \text{ false} \\ w_i\underline{e_j} & \delta_i(\underline{x}) = \text{ true and } x_i = v_{i,j}, \end{array} \right\}$$

where $\underline{e_j} \in \mathbb{R}^{m_i}$ is zero in all dimensions except j, where it it 1.

Proposition 3. *Embedding ${f_i}^{(cat)}$ is an isometric embedding of $(\mathcal{X}, {d_i}^{(cat)})$ into $(\mathbb{R}^{m_i}, {d_E}^{m_i})$.*

Proof. Consider two inputs $\underline{x}, \underline{x}' \in \mathcal{X}$. As in the proof of Proposition 1, we need to show that ${d_i}^{(cat)}(\underline{x}, \underline{x}') = {d_E}^{m_i}({f_i}^{(cat)}(\underline{x}), {f_i}^{(cat)}(\underline{x}'))$ and consider the following cases.

Case 1: $\delta_i(\underline{x}) = \delta_i(\underline{x}') =$ false. In this case, we trivially have

$${d_E}^{m_i}({f_i}^{(cont)}(\underline{x}), {f_i}^{(cont)}(\underline{x}')) = {d_E}^{m_i}(\underline{0}, \underline{0}) = 0 = {d_i}^{(cont)}(\underline{x}, \underline{x}').$$

Case 2: $\delta_i(\underline{x}) \neq \delta_i(\underline{x}')$. In this case, we have

$${d_E}^{m_i}({f_i}^{(cat)}(\underline{x}), {f_i}^{(cat)}(\underline{x}')) = {d_E}^{m_i}(w_i\underline{e_j}, \underline{0}) = w_i = d_i(\underline{x}, \underline{x}'),$$

and symmetrically for $d_E(\underline{0}, w_i\underline{e_j})$.

Case 3: $\delta_i(\underline{x}) = \delta_i(\underline{x}') =$ true. If $x_i = x_i' = v_{i,j}$, we have

$${d_E}^{m_i}({f_i}^{(cat)}(\underline{x}), {f_i}^{(cat)}(\underline{x}')) = {d_E}^{m_i}(w_i\underline{e_j}, w_i\underline{e_j}) = 0 = {d_i}^{(cat)}(\underline{x}, \underline{x}')$$

If $x_i = v_{i,j} \neq v_{i,j'} = x'_i =$, we have

$$d_E({f_i}^{(cat)}(\underline{x}), {f_i}^{(cat)}(\underline{x}')) \quad = \quad {d_E}^{m_i}(w_i \underline{e_j}, w_i \underline{e_{j'}}) = w_i \sqrt{2} = {d_i}^{(cat)}(\underline{x}, \underline{x}')$$

□

Proposition 4. $\kappa\big({d_i}^{(cat)}(\cdot,\cdot)\big)$ *is a positive semi-definite covariance function over input space* $\mathcal{X}$.

Proof. This is a trivial extension to the proof of Proposition 2. □

FH: missing - the trivial part where we multiply together all the kernels to define a PSD kernel for the whole space.