today: continue RKHS

**representer's thm:** says that (for $H$ a RKHS) $f^*$

$$
\min_{f \in H} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \langle f, f \rangle_H
$$

is reached for $f^* = \sum_{i=1}^{n} \alpha_i^* k(x_i, x)$.

$$
\alpha^* = \frac{1}{\langle k(\cdot, x), k(\cdot, x) \rangle_{H}} \langle k(\cdot, x), y \rangle
$$

If $f^* = \sum_{i=1}^{n} \alpha_i^* k(x_i, x)$, then $\|f^*\|^2_{H} = \langle f^*, f^* \rangle_{H} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j) = \alpha^T K \alpha$

$(K)_{ij} = k(x_i, x_j)$ (n x n matrix)

Let inner product of $(f(x_i))$ on data

$$
\min_{\alpha \in \mathbb{C}^n} \frac{1}{2} \sum_{i=1}^{n} (y_i - \langle \alpha, k(x_i, \cdot) \rangle)^2 + \alpha^T K \alpha
$$

finale: dim opt.

**getting a handle on $H$:**

- Generalize diagonalization of matrices to semi-definite $K$ is in $H$:
  - $K$ is a valid kernel $\Rightarrow K \geq 0$
$K$ is a valid kernel $\iff K \succeq 0$

$H = \text{span} \{ k(x_i, \cdot) : i = 1, \ldots, n \}$

$\{ k(x_i, \cdot) : \alpha \in \mathbb{R}^n \}$ is $1 \times n$ matrix

$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} \in \mathbb{R}^n$

$K \succeq 0 \iff 
K = U \Lambda U^T$ with

$\Lambda_{ii} \geq 0 \forall i$.

and $U$ is an orthonormal basis of $\mathbb{R}^n$

i.e. $U = (\psi_1, \ldots, \psi_n)$

$U^T U = I_n$ i.e. $\langle \psi_i, \psi_j \rangle = \delta_{ij}$

we can let

$\Phi = \sqrt{\Lambda} U^T$

$\Rightarrow K = \Phi^T \Phi$

$\Phi = \begin{bmatrix} \sqrt{\lambda_1} \psi_1^T \\ \vdots \\ \sqrt{\lambda_n} \psi_n^T \end{bmatrix}$

$\Phi^T \Phi = 
\begin{bmatrix} \lambda_1 \langle \psi_1, \psi_1 \rangle & \cdots & \lambda_n \langle \psi_1, \psi_n \rangle \\ \vdots & \ddots & \vdots \\ \lambda_n \langle \psi_n, \psi_1 \rangle & \cdots & \lambda_n \langle \psi_n, \psi_n \rangle \end{bmatrix}$

$k(x_i, x_j) = \lambda_i \langle \psi_i, \psi_j \rangle$

$\langle \Phi(x_i), \Phi(x_j) \rangle = \lambda_i \delta_{ij}$

Note: $K \psi_j = \lambda_j \psi_j$

$\Rightarrow k(x_i, x_j) = \langle \psi_i, \psi_j \rangle = \delta_{ij}$

back to SVM view: $H \subseteq \mathbb{R}_+$; $\forall \in H \Rightarrow v = Kx'$ for some $x' \in \mathbb{R}^n$

to get $\|v\|_H$, we compute $\alpha^*_v = K_{+v}^T$ pseudo-inverse

IFT6132 Page 2
to get \( \|v\|_H \), we compute \( \alpha_v = k^T v \)

\[
K = U L U^T
\]

\[
K^T = U L^T U^T
\]

so

\[
\|v\|_H^2 = \alpha_v^T K \alpha_v = v^T k^T k v \]

\[
= v^T U L U^T U (I_H) U^T v = v^T k^T v
\]

\[
= (U v)^T L^T (U v) \implies U v \text{ is projection of } v \text{ on } \mathbb{R}^n_{\text{basis}}
\]

so let \( v = \sum_{j=1}^{n} \beta_j y_j \) i.e. \( \beta_j = \langle v, y_j \rangle \_2 \)

\[
\|v\|_H^2 = \sum_{j=1}^{n} \beta_j^2
\]

\[
\|v\|_2^2 = \sum_{j=1}^{n} \beta_j^2
\]

and \( \|v\|_H^2 = \frac{1}{\lambda_2} \sum_{j=1}^{n} \beta_j^2 \)

so orthonormal basis of \( H \) in \( \mathbb{R}^2 \) is \( \sum_{j=1}^{n} \beta_j y_j \_2 \)

\( \|v\|_H \)

\[
\text{generalization to } \mathbb{R}^d \text{ dim space}.
\]

suppose \( X \) is a compact space + Lebesgue measure (e.g. \( X = [0,1] \))

\[
\mathcal{L}_2(X) = \sum_{f \in X} \|f\|_2^2 = \int_X \|f(x)\|^2 dx < \infty
\]
Let \( K \) be a continuous psd kernel \( f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \) with respect to standard norm on \( \mathbb{R} \). 

Define \( L_k : \mathbb{L}_2 \rightarrow \mathbb{L}_2 \) 

\[ \mathbb{L}_2 = \{ f \in \mathbb{C}(\mathbb{R}) \mid \int_{\mathbb{R}} |f(x)|^2 \, dx < \infty \} \]

\[ \text{such that} \quad \sum_{n} \mathcal{L}_k f_i(\cdot,x) \mathcal{L}_k f_i(\cdot,x) \, dx \]

then can show that \( L_k \) is a compact self-adjoint positive operator 
and yields an orthonormal basis \((\mathcal{L}_k f_i)_{i=1}^{\infty}\) of \( \mathbb{L}_2 \) with non-negative eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq 0 \)

\[ \text{i.e.} \quad L_k f_i = \lambda_i f_i \]

and we have 

\[ \mathcal{L}_k f(x) = \sum_{i=1}^{\infty} \lambda_i f_i(x) \, \mathcal{L}_k f_i(x) \] 

\[ \text{take} \quad K = \Phi \Phi^T \quad \text{of before} \]

\[ \text{a) feature space} \quad \mathcal{H} \subseteq \mathbb{L}_2 \quad \text{of} \quad \Phi : \mathbb{R} \rightarrow \mathbb{L}_2 \quad \text{with} \quad (\Phi(x))_i = \sqrt{\lambda_i} \, f_i(x) \]

\[ \text{here} \quad \mathcal{L}_k f(x) = \langle \Phi(x), \Phi(x) \rangle_{\mathbb{L}_2} \]

\[ \text{"diagonalize representation"}\]

\[ \text{here identify} \quad \mathcal{L}_k f(x) \in \mathbb{L}_2 \quad \text{as} \quad \Phi(x) \in \mathbb{L}_2 \]

\[ \text{(do not what} \quad \Phi \in \mathbb{L}_2 \quad \text{looks though)} \]

\[ \text{b) view} \quad \mathcal{H} \subseteq \mathbb{L}_2 \quad \text{as} \quad \mathcal{H} \subseteq \mathbb{L}_2 \quad \text{ellipsoid in} \quad \mathbb{L}_2 \]
b) $\ell_2$ norm

$$H \subseteq S_2 : H = \{ f \in S_2 : \sum_{i=1}^{d} \left( f_i^2 \right) \leq \lambda \}$$

and $\langle f, g \rangle_H = \sum_{i=1}^{d} \frac{f_i g_i}{\lambda}$

If $k$ is "universal"

$\Rightarrow$ $H_k$ is dense in $S_2$, i.e., for any $f \in S_2$

$\exists$ sequence $h_n \in H_{st.}$ s.t. $\| f - h_n \|_H \to 0$

Note: If $f \not\in H$; $\Rightarrow$ $\| f \|_H \to \infty$

× non-parametric learning:

$$\hat{f}_n = \operatorname{argmin}_{f \in H} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda_n \| f \|_H^2$$

$\hat{f}_n \to \operatorname{argmin}_{f \in H} \mathbb{E} S(f, X)$ (perhaps $\hat{f}_n \not\in H$

but $H$ dense in $S_2$

+ $\lambda_n$ properly of $\| f \|_H$ + consistency of $\hat{f}_n$ at correct rate.

Example: SVM with RBF kernel is "universally consistent" when $\lambda_n \to 0$.