Conditional Embeddings

For certain \( A \), can reconstruct \( p \) from the summary!

\[
\begin{align*}
\mu_Y &= \mathbb{E}_Y [Y] \\
\sigma_Y^2 &= \mathbb{V}_Y [Y] \\
\rho_Y &= \frac{\mathbb{C}_Y}{\sigma_Y \sigma_Y} \quad \text{(expected feature)}
\end{align*}
\]

Conditional Embedding

\[
U_{Y|X} := C_Y X C_X^{-1}
\]

- Kernel estimate (m-by-m matrix)
- \( \Phi(K + \lambda I)^{-1}Y^T \)

Key Idea

- If \( x_1 \) and \( x_2 \) are close, \( U_{Y|x_1} \) and \( U_{Y|x_2} \) will be similar.

Generalize to unobserved \( x \) via a kernel on \( X \)

Sum and Product Rules

<table>
<thead>
<tr>
<th>Probabilistic Relation</th>
<th>Hilbert Space Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_X = \int \mu_Y \pi(Y</td>
<td>X) dY )</td>
</tr>
<tr>
<td>( \mu_{XY} = \int \mu_{YX} \pi(Y</td>
<td>X) dY )</td>
</tr>
</tbody>
</table>

Avantages

- Applies to any domain with a kernel
  - Eg. reals, rotations, permutations, strings
- Applies to more general distributions
  - Eg. multimodal, skewed

Future Directions

- Learn dynamical systems from sequences of observations (without the hidden states)
- Generalize inference procedures to handle more complicated models (e.g., Bayesian networks, Markov random fields)
- Formulate conditional independence tests
- Establish sample complexity bounds for the dynamical systems setting

Conclusion

- Embedding conditional distributions
- Generalizes to unobserved \( X \) using kernels
- Theoretical analysis of empirical estimator convergence
- Application to dynamical systems learning and inference