Nonlinear Transformation (LFD 3.4)
Linear Hypotheses

- Up to now: linear hypotheses
  - Perceptron, Linear regression, Logistic regression, ···
- Many problems are not linearly separable
Circular Separable

\[ D \text{ is not linear separable} \]
Circular Separable

$\mathcal{D}$ is not linear separable but circular separable by a circle of radius $\sqrt{0.6}$ centered at origin:

$$h_{\text{SEP}}(x) = \text{sign}(-x_1^2 - x_2^2 + 0.6)$$
Circular Separable

$\mathcal{D}$ is not linear separable

but **circular separable** by a circle of radius $\sqrt{0.6}$ centered at origin:

$$h_{\text{SEP}}(\mathbf{x}) = \text{sign}(-x_1^2 - x_2^2 + 0.6)$$

How to extend perceptron to this case?
Circular Separable and Linear Separable

\[ h(x) = \text{sign}(0.6 \cdot 1 + (-1) \cdot x_1^2 + (-1) \cdot x_2^2) \]
Circular Separable and Linear Separable

\[
\begin{align*}
\tilde{h}(x) &= \text{sign}\left(\sum_{i=0}^{2} \tilde{w}_i \cdot x_i \right) \\
&= \text{sign}(\tilde{\mathbf{w}}^T \mathbf{z})
\end{align*}
\]
Circular Separable and Linear Separable

\[ h(x) = \text{sign}(0.6 \cdot \tilde{w}_0 \cdot z_0 + (-1) \cdot \tilde{w}_1 \cdot x_1^2 + (-1) \cdot \tilde{w}_2 \cdot x_2^2) \]

\[ = \text{sign}(\tilde{w}^T z) \]

\[ \{(x_n, y_n)\} \text{ circular separable} \Rightarrow \{(z_n, y_n)\} \text{ linear separable} \]
Circular Separable and Linear Separable

\[ h(x) = \text{sign} \left( \frac{0.6}{\tilde{w}_0} \cdot 1 + \frac{-1}{\tilde{w}_1} \cdot x_1^2 + \frac{-1}{\tilde{w}_2} \cdot x_2^2 \right) \]

\[ = \text{sign}(\tilde{w}^T z) \]

\( \{(x_n, y_n)\} \) circular separable \( \Rightarrow \{(z_n, y_n)\} \) linear separable

\( x \in \mathcal{X} \rightarrow z \in \mathcal{Z} \) (using a nonlinear transformation \( \Phi \))
Nonlinear Transformation

- Define nonlinear transformation

\[ \Phi(x) = (1, x_1^2, x_2^2) = (z_0, z_1, z_2) = z \]
Nonlinear Transformation

- Define nonlinear transformation

\[ \Phi(x) = (1, x_1^2, x_2^2) = (z_0, z_1, z_2) = z \]

- Linear hypotheses in \( Z \) space:

\[ \text{sign}(\tilde{h}(z)) = \text{sign}(\tilde{h}(\Phi(x))) = \text{sign}(w^T \Phi(x)) \]
Nonlinear Transformation

- Define nonlinear transformation

\[ \Phi(x) = (1, x_1^2, x_2^2) = (z_0, z_1, z_2) = \mathbf{z} \]

- Linear hypotheses in \( \mathcal{Z} \) space:

\[ \text{sign}(\tilde{h}(\mathbf{z})) = \text{sign}(\tilde{h}(\Phi(x))) = \text{sign}(\mathbf{w}^T \Phi(x)) \]

- Lines in \( \mathcal{Z} \) space \( \Leftrightarrow \) some quadratic curves in \( \mathcal{X} \)-space
A “bigger” $\mathcal{Z}$-space:

$$\Phi_2(x) = (1, x_1, x_2, x_1^2, x_1 x_2, x_2^2)$$
A “bigger” $\mathcal{Z}$-space:

$$\Phi_2(x) = (1, x_1, x_2, x_1^2, x_1 x_2, x_2^2)$$

Perceptrons in $\mathcal{Z}$-space $\iff$ quadratic hypotheses in $\mathcal{X}$-space
General Quadratic Hypothesis Set

- A “bigger” \( \mathcal{Z} \)-space:
  \[
  \Phi_2(x) = (1, x_1, x_2, x_1^2, x_1x_2, x_2^2)
  \]

- Perceptrons in \( \mathcal{Z} \)-space \( \iff \) quadratic hypotheses in \( \mathcal{X} \)-space

- The hypotheses space:
  \[
  \mathcal{H}_{\Phi_2} = \{ h(x) : h(x) = \tilde{w}^T \Phi_2(x) \text{ for some } \tilde{w} \}
  \]
  (Quadratic hypotheses)
General Quadratic Hypothesis Set

- A “bigger” \( Z \)-space:

\[
\Phi_2(x) = (1, x_1, x_2, x_1^2, x_1x_2, x_2^2)
\]

- Perceptrons in \( Z \)-space \( \iff \) quadratic hypotheses in \( X \)-space

- The hypotheses space:

\[
H_{\Phi_2} = \{ h(x) : h(x) = \tilde{w}^T \Phi_2(x) \text{ for some } \tilde{w} \}
\]

(Quadratic hypotheses)

- Also include lines as degenerate cases
Learning a good quadratic function

- Transform original data \( \{x_n, y_n\} \) to \( \{z_n = \Phi(x_n), y_n\} \)
- Solve a linear problem on \( \{z_n, y_n\} \) using your favorite algorithm \( A \) to get a good model \( \tilde{w} \)
- Return the model \( h(x) = \text{sign}(\tilde{w}^T \Phi(x)) \)
Polynomial mappings

- Can now freely do quadratic PLA, quadratic regression, …
Polynomial mappings

- Can now freely do quadratic PLA, quadratic regression, ⋅⋅⋅
- Can easily extend to any degree of polynomial mappings
  
  E.g., \( \Phi(x) = (x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1x_2^2, x_1x_3^2, x_1^2x_3, x_2^2x_3, x_1^2x_3, x_2^2x_3, x_1^3, x_2^3, x_3^3) \)
The price we pay

- $Q$-th order polynomial transform:
  
  $$\Phi(x) = (1, x_1, x_2, \cdots, x_d, x_1^2, x_1 x_2, \cdots, x_2^2, \cdots, x_d^Q, x_1^{Q-1} x_2, \cdots, x_d^Q)$$

- What’s the dimension?
The price we pay

- $Q$-th order polynomial transform:

$$
\Phi(x) = (1, x_1, x_2, \cdots, x_d, \\
x_1^2, x_1 x_2, \cdots, x_d^2, \\
\cdots \\
x_1^Q, x_1^{Q-1} x_2, \cdots, x_d^Q)
$$

- What’s the dimension?

Number of ways of $\leq Q$-combination from $d$ kinds (with repetitions) = $O(d^Q)$
The price we pay

- $Q$-th order polynomial transform:

$$
\Phi(x) = (1, x_1, x_2, \ldots, x_d,
\quad x_1^2, x_1 x_2, \ldots, x_2^2, \ldots
\quad x_1^Q, x_1^{Q-1} x_2, \ldots, x_d^Q )
$$

- What’s the dimension?
  Number of ways of $\leq Q$-combination from $d$ kinds (with repetitions)
  $= O(d^Q)$

- Need to compute and store $z = \Phi(x)$ and $\tilde{w}$
  $O(d^Q)$ dimensional vectors
Model Complexity Price

- $\tilde{w}$: $O(d^Q)$-dimensional vector
- VC-dimension: $d_{VC} \approx O(d^Q)$
Model Complexity Price

- $\tilde{w}$: $O(d^Q)$-dimensional vector
- VC-dimension: $d_{VC} \approx O(d^Q)$
- $Q$ increase (higher order polynomial)
  $\Rightarrow$ smaller $E_{in}$
  $\Rightarrow$ Larger gap between $E_{in}$ and $E_{out}$
Conclusions

- Next class: LFD 4.1

Questions?