

8.

9. (a) let $K_c(\mathbf{x}, \mathbf{z}) = c + \sum_{i=1}^n x_i z_i = c + \mathbf{x}^T \mathbf{z}$

Our function $K_c(\cdot, \cdot)$ is psd iff:

$\forall \{u_i\}_{1:m} \in \mathcal{R}, \{\mathbf{x}_j\}_{1:m} \in \mathcal{R}^n$:

$\sum_{k,l} u_k u_l K_c(\mathbf{x}_u, \mathbf{x}_l) \geq 0$ We expand our kernel function to give:

$$\sum_{k,l} u_k u_l K_c(\mathbf{x}_u, \mathbf{x}_l) = \sum_{n,m} c u_n u_m + \sum_{n,m} \langle \mathbf{x}_n, \mathbf{x}_m \rangle u_n u_m$$

Here, we note that $\langle \cdot, \cdot \rangle$ is an inner product over the Hilbert space \mathcal{R}^n and hence is psd by the representer theorem:

$$\implies \sum_{n,m} \langle \mathbf{x}_n, \mathbf{x}_m \rangle u_n u_m \geq 0 \quad (1)$$

Proposition

$K_c(\cdot, \cdot)$ is psd iff $c \geq 0$

Proof:

(\implies)

Suppose $c \geq 0$.

$$(1) \implies \sum_{k,l} u_k u_l K_c(\mathbf{x}_u, \mathbf{x}_l) \geq c \sum_{n,m} u_n u_m = c (\sum_n u_n)^2 \geq 0$$

(\impliedby)

let $c < 0$. Suppose our vectors x_n are identically equal to $(\sqrt{a}, 0)^T$, for some $a < |c|/n \implies x_k^T x_l = a$

$\forall k, l$.

$$\implies \sum_{k,l} u_k u_l K_c(\mathbf{x}_u, \mathbf{x}_l) = (a + c) \sum_{n,m} u_n u_m = (a + c) (\sum_n u_n)^2 < 0$$

Hence for any $c < 0$, $\exists \{u_i\}_{1:m} \in \mathcal{R}, \{\mathbf{x}_j\}_{1:m} \in \mathcal{R}^n$ s.t:

$$\sum_{k,l} u_k u_l K_c(\mathbf{x}_u, \mathbf{x}_l) < 0 \square$$

(b) Using this kernel in our ridge regression, we arrive at the following prediction function:

$$\hat{y}_{test} = lc + \sum_i \alpha_i x_i \cdot x_{test}$$

where $\alpha = (K + \gamma I_l + c 1_{n \times n})^{-1} y_{train}$

Where $1_{n \times n}$ is a matrix of all ones.

We note that as c becomes large, $\alpha \rightarrow 0$ and hence the predictions for \hat{y}_{test} approach the constant function $f(x) = lc$

- (c) Define $g(t) := \text{sgn}(f(t)) = \frac{f(t)}{|f(t)|}$

Proposition:

as $\beta \rightarrow \infty$, $g(t) \rightarrow 1 - NN$

Proof:

$$f(t) = \sum_i \alpha_i K(x_i, t), \text{ where } \alpha = (K + \gamma I_l)^{-1} y$$

since $K(x, x) = \exp(0) = 1$ for all x , we decompose K as follows:

$$K = I_l + \tilde{K}, \text{ where } \tilde{K}_{ij} = K(x_i, x_j), i \neq j. \text{ We note here that as } \beta \rightarrow \infty, \tilde{K} \rightarrow 0.$$

We take a Taylor expansion in \tilde{K} , to arrive at the following result:

$$\alpha = ((\gamma + 1)I_l + \tilde{K})^{-1} y = \frac{1}{\gamma + 1} (I_l - O(\tilde{K})) y \text{ Hence as } \beta \rightarrow \infty, \alpha \rightarrow \frac{1}{\gamma + 1} y$$

Hence our predictor function becomes:

$$f(t) \rightarrow \frac{1}{\gamma + 1} \sum_i y_i K(x_i, t) \implies g(t) \rightarrow \frac{\sum_i y_i K(x_i, t)}{|\sum_i y_i K(x_i, t)|}$$

Let's assume that there exists a unique point x in our training set of minimal distance to our test point.

Define $d_i := \|x_{test} - x_i\|^2$, and $d_* := \min_i d_i$

$$g(t) = \frac{y_* \exp(-\beta d_*)}{|\sum_i y_i \exp(-\beta d_i)|} + \sum_{i \neq *} \frac{y_i \exp(-\beta d_i)}{|\sum_i y_i \exp(-\beta d_i)|} = \frac{y_*}{|y_* + \sum_{i \neq *} y_i \exp(-\beta(d_i - d_*))|} + \sum_{i \neq *} \frac{y_i \exp(-\beta(d_i - d_*))}{|y_* + \sum_{i \neq *} y_i \exp(-\beta(d_i - d_*))|}.$$

Note for $d_i \neq d_*$, $\exp(-\beta(d_i - d_*)) \rightarrow 0$ as $\beta \rightarrow \infty$. Hence:

$$g(t) \rightarrow \frac{y_*}{|y_* + 0|} + \frac{0}{|y_* + 0|} = y_* = y_i : x_i = \min_i \|x_{test} - x_i\| \text{ which is the required 1-NN predictor.}$$

□

- (d) Given a whack-a-mole grid of size n , it is solvable if we can apply a sequence from our n^2 permissible moves that results in a blank screen. We first index all permissible moves by $T_{i,j}$, which has the effect of flipping the state of squares (i, j) , $(i - 1, j)$, $(i + 1, j)$, $(i, j - 1)$, $(i, j + 1)$, as long as each coordinate is still on the grid.

We may consider each grid as an element of the space of square matrices over \mathcal{F}_2 , the finite field of order 2: under this construction, we see that the application of a move $T_{i,j}$ corresponds to the sum between the grid and a matrix T_{ij} , where T_{ij} , where the entries of T_{ij} are zero everywhere except at (i, j) , $(i - 1, j)$, $(i + 1, j)$, $(i, j - 1)$, $(i, j + 1)$, if these values are present on the grid. For example, if $n = 4$, the transformation $T_{2,2}$ corresponds to the addition of the matrix T_{22} :

$$T_{22} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Since our operations on a grid are now simply the sum of matrices over a \mathcal{F}_2 , it is now immediately obvious that our operations are commutative, by the commutativity of addition.

Further, since applying the same operation twice would be the same as applying the matrix twice in order, we can see that any operation applied twice would have the effect of applying no operation at all (since $T_{ij} + T_{ij} = 0$).

Hence, we conclude for $1 \leq i, j \leq n$, each operation is applied at most once. From this point, we note that there is no reason for our operations to be square matrices, as the only required properties of the space are those of a vector space (we make no use here of matrix multiplication). Hence, we may vectorise both our input grid and our operation matrices to get a set of vectors over $F_2^{n^2}$. We now index our operation vectors $t_i : i = 1, \dots, n^2$.

Let $x_i = \mathcal{I}[t_i \text{ is used}]$ be an indicator function representing whether t_i is added to the input to get our blank grid.

We observe that our problem is solvable for a given input grid G (now vectorised as a vector \mathbf{g}) iff there exists a vector $\mathbf{x} \in F_2^{n^2}$ st:

$$\sum_i t_i x_i + \mathbf{g} = \mathbf{0}.$$

If we stack the vectors t_i horizontally as the columns of a matrix T , we get:

$$T\mathbf{x} + \mathbf{g} = \mathbf{0}$$

Since every element is its own additive inverse over F_2 , we add \mathbf{g} to both sides and arrive at the linear system $T\mathbf{x} = \mathbf{g}$.

Since the matrix T may not be invertible in general, we perform gaussian elimination over F_2 , i.e create an augmented matrix $[T|\mathbf{g}]$ and reduce to row echelon form. If the problem is solvable, we will end up with a consistent (but perhaps underdetermined) reduced system of equations. If the input has no solution, the reduced system of equations will have some inconsistency in its variables: there will be some row of the reduced row echelon augmented matrix with all zeros to the left of the augmentation line, and a non-zero entry to the right of the augmentation line.

Finally, we make a note of the time complexity of this algorithm: the creation of our matrix T can be performed in $O(n^4)$ since we have n^2 operation grids of size n^2 . Further, gaussian elimination can be performed in $O(n^3)$ operations in any field, and hence our total running time is $O(n^3)$.

Once we have determined whether a grid configuration is solvable, it remains to find a valid solution. We note that if a solution exists, then a least squares solver should find it. This operation should take $O(n^3)$ operations.

1.