Exercises for “Vertiefung Neuronale Netze”
SS 2018 Sheet 8

Due on: 18.5.2018

Task 8.1, Projection:

1. Let u be a normalized vector ($\|u\| = 1$). Show that the matrix $P = uu^t$ is a projection matrix (i.e. $P^2 = P$) and projecting to the vector u, i.e. $Pu = u$.
2. Show that the matrix $P_\perp = \mathbf{1} - uu^t$ is the projection onto the orthogonal space of u, i.e. $(P_\perp v)^t \cdot u = 0$ for any vector v.

For the following tasks, use `matlab` to analyse the properties of Partial-Least-Squares-Regression. Consider a training set comprising M samples of dimensionality d. Only two of these dimensions are relevant (with different variances σ_1^2, σ_2^2) while $d-2$ are irrelevant input dimensions (with uniform variance σ_d^2). The output y is linearly dependent on the two relevant dimensions like so:

$$y = w_1 * x_1 + w_2 * x_2 \qquad w_1 = 4 \text{ and } w_2 = 1 \tag{1}$$

Task 8.2, : What would be the ideal result of PLS regression?

Task 8.3, : Perform PLS manually:

- Create a data matrix $\mathbf{X} \in \mathbb{R}^{M \times d}$ of uniformly distributed random variables in the interval $[-1, 1]$:
  `X = -1 + 2*rand(M, d);`
- Generate the associated output matrix $\mathbf{Y} \in \mathbb{R}^{M \times 1}$ (with $w_1 = 4, w_2 = 1$):
  `f=@(x,y)(4*x+y); Y=arrayfun(f, X(:,1), X(:,2));`
- Visualize the output as a function of x_1, x_2:
  `plot3(X(:,1),X(:,2),Y,'.');`
  or as a surface mesh:
  `U=-1:0.1:1; U=repmat(U,length(U),1); V=U'; W=arrayfun(f,U,V); surf(U,V,W);`
- Compute the first two PLS projection directions u_i and the corresponding regression weights β_i. Compare them with the gradient of the function. How many projection directions are required?
- How does the function approximation looks like considering nonlinear functions e.g. $y = x_1 * x_2$ or $y = x_1 + 4 * sin(2 * x_2)$? How many projection directions are needed now?

Task 8.4, : Compare the results from PLS under the following conditions:

- small variance for irrelevant inputs: $\sigma_1 = \sigma_2 = 1$ und $\sigma_d = 0.1$
- different variance in the relevant inputs: $\sigma_1 = 10, \sigma_2 = 1$ und $\sigma_d = 0.1$
- equal variance for the irrelevant inputs: $\sigma_1 = \sigma_2 = \sigma_d = 1$
- higher variance for the irrelevant inputs: $\sigma_1 = \sigma_2 = 1$ und $\sigma_d = 10$

How can you interpret the results? How does the influence of the irrelevant inputs change with varying size of the dataset $M \in \{100, 500, 5000\}$?