Zero-Shot Kernel Learning

Supplementary Material

Hongguang Zhang*,2,1 Piotr Koniusz*,1,2
[1]Data61/CSIRO, [2]Australian National University
firstname.lastname@{data61.csiro.au[1], anu.edu.au[2]}

1. Weak incoherence of $\boldsymbol{W}$

Gaussian-Ort									
1.0	0.0384	-0.0368	-0.094	-0.1192	-0.2234	-0.0582	-0.1484	0.0876	0.0628
0.0384	1.0	0.0075	-0.2262	-0.0523	-0.0172	-0.0501	-0.0689	0.0674	0.1292
-0.0368	0.0075	1.0	-0.0715	0.1992	0.0166	0.0715	0.0342	-0.1668	0.0438
-0.094	-0.2262	-0.0715	1.0	0.0912	-0.0582	-0.053	0.0621	-0.0905	-0.1422
-0.1192	-0.0523	0.1992	0.0912	1.0	-0.1231	-0.0914	-0.1321	0.0249	-0.065
-0.2234	-0.0172	0.0166	-0.0582	-0.1231	1.0	0.0539	-0.0298	-0.0015	0.074
-0.0582	-0.0501	0.0715	-0.053	-0.0914	0.0539	1.0	-0.1084	-0.0629	0.0396
-0.1484	-0.0689	0.0342	0.0621	-0.1321	-0.0298	-0.1084	1.0	0.0321	0.3139
0.0876	0.0674	-0.1668	-0.0905	0.0249	-0.0015	-0.0629	0.0321	1.0	0.029
0.0628	0.1292	0.0438	-0.1422	-0.065	0.074	0.0396	0.3139	0.029	1.0

Table 1: Illustration of the first 10×10 elements of $\boldsymbol{W}^T\boldsymbol{W}$ for (*Gaussian-Ort*). We ℓ_2-norm normalized columns of $\boldsymbol{W}$ and then color-coded cells. The intense red indicates closeness of the off-diagonal values to one. As can be seen, the off-diagonal entries have much lower values than the elements on the diagonal.

Gaussian									
1.0	0.259	-0.0556	-0.172	-0.1865	-0.1168	-0.1503	-0.1406	0.2686	0.2076
0.259	1.0	0.0257	-0.4768	-0.1866	-0.096	-0.1888	-0.0923	0.3299	0.2592
-0.0556	0.0257	1.0	-0.3542	0.2656	-0.1339	-0.0388	-0.0986	-0.1592	-0.1455
-0.172	-0.4768	-0.3542	1.0	-0.1488	-0.0246	0.048	-0.0241	0.0021	-0.0497
-0.1865	-0.1866	0.2656	-0.1488	1.0	-0.2122	-0.0162	-0.2212	-0.1753	-0.2876
-0.1168	-0.096	-0.1339	-0.0246	-0.2122	1.0	0.6296	0.4083	-0.057	-0.0863
-0.1503	-0.1888	-0.0388	0.048	-0.0162	0.6296	1.0	-0.0041	-0.0791	-0.0859
-0.1406	-0.0923	-0.0986	-0.0241	-0.2212	0.4083	-0.0041	1.0	0.0188	0.1778
0.2686	0.3299	-0.1592	0.0021	-0.1753	-0.057	-0.0791	0.0188	1.0	0.3737
0.2076	0.2592	-0.1455	-0.0497	-0.2876	-0.0863	-0.0859	0.1778	0.3737	1.0

Table 2: Illustration of the first 10×10 elements of $\boldsymbol{W}^T\boldsymbol{W}$ for (*Gaussian*). We ℓ_2-norm normalized columns of $\boldsymbol{W}$ and then color-coded cells. The intense red indicates closeness of the off-diagonal values to one. As can be seen, the off-diagonal entries have values that are sometimes comparable to the values on the diagonal.

To demonstrate that our solution learns $\boldsymbol{W}$ with weakly incoherent w.r.t. each other column vectors $\boldsymbol{w}_1, ..., \boldsymbol{w}_{d'}$ for (*Gaussian-Ort*), we display the first 10×10 entries of $\boldsymbol{W}^T\boldsymbol{W}$ for $\boldsymbol{W}$ with ℓ_2-norm normalized column vectors. We learned $\boldsymbol{W}$ on the AWA1 dataset. As can be seen from Table 1, the values of the off-diagonal elements are much lower than the values of the elements on the diagonal. This result implies the weak incoherence between the column vectors of $\boldsymbol{W}$. In contrast, Table 2 presents the first 10×10 entries of $\boldsymbol{W}^T\boldsymbol{W}$, where $\boldsymbol{W}$ contains ℓ_2-norm normalized column vectors, for (*Gaussian*) which does not impose the soft/implicit incoherence constraints. We display results for $\boldsymbol{W}$ learned on the AWA1 dataset. As can be seen from the table, the values of the off-diagonal elements are often comparable to the values of the elements on the diagonal. This result implies there is no incoherence between the column vectors of $\boldsymbol{W}$.

*Both authors contributed equally.