Supplement to the paper 'A proposal for PU classification under Non-SCAR using clustering and logistic model'

Konrad Furmańczyk and Kacper Paczutkowsk

Institute of Information Technology, Warsaw University of Life Sciences, Warsaw, Poland

This supplement contains 3 sections. In Section 1, tables for all classification metrics across 12 datasets and figures for the non-SCAR scheme with the executing time of considered algorithms are presented. Analogous results for the SCAR scheme are presented in Section 2. Section 3 contains the description of the non-SCAR labeling function. In all tables, we reported mean and standard deviations (in brackets). The best results are marked.

1 Non-SCAR Scheme

1.1 Tables for all classification metrics for 12 datasets.

q	naive acc	clust acc	strict lassclust acc	non strict lassclust acc	lasso joint acc	naive f1	clust f1	strict lassclust f1	non strict lassclust f1	lasso joint f1	naive auc	clust auc	strict lassclust auc	non strict lassclust auc	lasso joint auc
0.25	**0.64 (0.16)**	**0.71 (0.16)**	0.7 (0.15)	**0.71 (0.16)**	**0.53 (0.15)**	**0.63 (0.18)**	0.54 (0.25)	0.5 (0.25)	0.54 (0.25)	0.56 (0.14)	**0.77 (0.13)**	**0.73 (0.17)**	**0.75 (0.16)**	**0.75 (0.16)**	**0.79 (0.15)**
0.5	**0.64 (0.16)**	**0.71 (0.16)**	0.69 (0.15)	**0.71 (0.16)**	**0.53 (0.15)**	**0.63 (0.18)**	0.54 (0.24)	0.52 (0.23)	0.54 (0.25)	0.56 (0.14)	**0.77 (0.13)**	**0.73 (0.17)**	**0.75 (0.16)**	**0.75 (0.16)**	0.78 (0.15)
1	**0.64 (0.16)**	**0.71 (0.16)**	**0.71 (0.16)**	**0.71 (0.16)**	**0.53 (0.15)**	**0.63 (0.18)**	**0.56 (0.24)**	**0.56 (0.24)**	**0.56 (0.24)**	**0.57 (0.15)**	**0.77 (0.13)**	**0.73 (0.17)**	**0.75 (0.16)**	**0.75 (0.16)**	**0.79 (0.15)**

Table 1. Summary of Mean and Standard Deviation for Accuracy, F1, and AUC Metrics by q (Highest Mean in Bold); non-SCAR scheme

Table 2. Summary Statistics for df = adult

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	**0.697 (0.006)**	0.452 (0.005)	**0.461 (0.006)**	**0.697 (0.006)**	0.452 (0.005)	**0.461 (0.006)**	**0.697 (0.006)**	0.452 (0.005)	0.461 (0.006)
0.3	clust	0.526 (0.010)	**0.625 (0.005)**	0.272 (0.005)	0.574 (0.009)	**0.644 (0.004)**	0.371 (0.014)	0.657 (0.004)	**0.647 (0.005)**	0.478 (0.005)
0.3	strict-lassclust	0.532 (0.010)	**0.625 (0.005)**	0.272 (0.005)	0.599 (0.009)	**0.644 (0.004)**	0.371 (0.013)	0.688 (0.006)	**0.647 (0.005)**	**0.479 (0.005)**
0.3	non-strict-lassclust	0.532 (0.010)	**0.625 (0.005)**	0.272 (0.005)	0.599 (0.009)	**0.644 (0.004)**	0.371 (0.013)	0.688 (0.006)	**0.647 (0.005)**	**0.479 (0.005)**
0.3	lassojoint	0.684 (0.009)	0.365 (0.055)	0.430 (0.020)	0.684 (0.009)	0.365 (0.055)	0.430 (0.020)	0.684 (0.009)	0.365 (0.055)	0.430 (0.020)
0.52	naive	0.702 (0.006)	0.570 (0.005)	0.490 (0.005)	0.702 (0.006)	0.570 (0.005)	0.490 (0.005)	0.702 (0.006)	0.570 (0.005)	0.490 (0.005)
0.52	clust	0.535 (0.008)	**0.629 (0.003)**	0.244 (0.007)	0.571 (0.008)	**0.636 (0.004)**	0.322 (0.007)	0.662 (0.004)	**0.645 (0.004)**	0.473 (0.006)
0.52	strict-lassclust	0.535 (0.008)	**0.629 (0.003)**	0.244 (0.007)	0.579 (0.008)	**0.636 (0.004)**	0.322 (0.007)	0.684 (0.005)	**0.645 (0.004)**	0.473 (0.006)
0.52	non-strict-lassclust	0.535 (0.008)	**0.629 (0.003)**	0.244 (0.007)	0.579 (0.008)	**0.636 (0.004)**	0.322 (0.007)	0.684 (0.005)	**0.645 (0.004)**	0.473 (0.006)
0.52	lassojoint	**0.731 (0.020)**	0.585 (0.007)	**0.500 (0.006)**	**0.731 (0.020)**	0.585 (0.007)	**0.500 (0.006)**	**0.731 (0.020)**	0.585 (0.007)	**0.500 (0.006)**
0.79	naive	**0.732 (0.002)**	**0.677 (0.001)**	0.484 (0.001)	**0.732 (0.002)**	**0.677 (0.001)**	0.484 (0.001)	**0.732 (0.002)**	0.677 (0.001)	0.484 (0.001)
0.79	clust	0.636 (0.002)	0.666 (0.001)	0.265 (0.001)	0.682 (0.001)	0.676 (0.001)	0.352 (0.003)	0.715 (0.001)	**0.685 (0.001)**	0.435 (0.001)
0.79	strict-lassclust	0.637 (0.002)	0.667 (0.001)	0.265 (0.001)	0.682 (0.001)	0.676 (0.001)	0.352 (0.002)	0.715 (0.002)	0.684 (0.001)	0.435 (0.001)
0.79	non-strict-lassclust	0.637 (0.002)	0.667 (0.001)	0.265 (0.001)	0.682 (0.001)	0.676 (0.001)	0.352 (0.002)	0.715 (0.002)	0.684 (0.001)	0.435 (0.001)
0.79	lassojoint	0.731 (0.002)	0.675 (0.005)	**0.488 (0.004)**	0.731 (0.002)	0.675 (0.005)	**0.488 (0.004)**	0.731 (0.002)	0.675 (0.005)	**0.488 (0.004)**

Table 3. Summary Statistics for df = artif

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.29	naive	**0.752 (0.019)**	**0.579 (0.016)**	**0.691 (0.015)**	**0.752 (0.019)**	**0.579 (0.016)**	**0.691 (0.015)**	**0.752 (0.019)**	**0.579 (0.016)**	**0.691 (0.015)**
0.29	clust	0.635 (0.084)	0.461 (0.102)	0.392 (0.102)	0.595 (0.063)	0.500 (0.076)	0.434 (0.075)	0.573 (0.051)	0.535 (0.054)	0.497 (0.058)
0.29	strict-lassclust	0.631 (0.080)	0.462 (0.096)	0.389 (0.126)	0.591 (0.061)	0.499 (0.070)	0.429 (0.116)	0.572 (0.051)	0.534 (0.053)	0.496 (0.058)
0.29	non-strict-lassclust	0.635 (0.085)	0.461 (0.099)	0.387 (0.096)	0.595 (0.064)	0.498 (0.076)	0.430 (0.070)	0.572 (0.051)	0.534 (0.053)	0.496 (0.058)
0.29	lassojoint	0.697 (0.033)	0.515 (0.039)	0.666 (0.018)	0.694 (0.031)	0.512 (0.038)	0.665 (0.018)	0.697 (0.031)	0.514 (0.042)	0.666 (0.019)
0.5	naive	0.793 (0.008)	0.637 (0.014)	0.717 (0.010)	**0.793 (0.008)**	0.637 (0.014)	**0.717 (0.010)**	**0.793 (0.008)**	0.637 (0.014)	0.717 (0.010)
0.5	clust	**0.868 (0.110)**	**0.757 (0.087)**	0.699 (0.123)	0.739 (0.131)	**0.669 (0.095)**	0.580 (0.101)	0.685 (0.019)	0.634 (0.022)	0.595 (0.027)
0.5	strict-lassclust	**0.868 (0.110)**	0.734 (0.064)	0.667 (0.057)	0.746 (0.131)	0.660 (0.084)	0.646 (0.136)	0.685 (0.020)	0.631 (0.021)	0.592 (0.027)
0.5	non-strict-lassclust	0.867 (0.111)	0.756 (0.090)	0.692 (0.139)	0.736 (0.130)	0.637 (0.078)	0.516 (0.061)	0.685 (0.020)	0.631 (0.021)	0.592 (0.027)
0.5	lassojoint	0.784 (0.010)	0.660 (0.014)	**0.728 (0.011)**	0.771 (0.017)	0.618 (0.036)	0.709 (0.021)	0.784 (0.010)	**0.660 (0.014)**	**0.728 (0.011)**
0.8	naive	**0.953 (0.000)**	**0.850 (0.000)**	**0.865 (0.000)**	**0.953 (0.000)**	**0.850 (0.000)**	**0.865 (0.000)**	**0.953 (0.000)**	**0.850 (0.000)**	**0.865 (0.000)**
0.8	clust	0.855 (0.000)	0.668 (0.000)	0.527 (0.000)	0.898 (0.000)	0.655 (0.000)	0.486 (0.000)	0.871 (0.000)	0.720 (0.000)	0.639 (0.000)
0.8	strict-lassclust	0.857 (0.000)	0.713 (0.000)	0.626 (0.000)	0.824 (0.000)	0.715 (0.000)	0.753 (0.000)	0.872 (0.000)	0.715 (0.000)	0.631 (0.000)
0.8	non-strict-lassclust	0.854 (0.000)	0.648 (0.000)	0.484 (0.000)	0.899 (0.000)	0.612 (0.000)	0.369 (0.000)	0.872 (0.000)	0.715 (0.000)	0.631 (0.000)
0.8	lassojoint	0.874 (0.000)	0.520 (0.000)	0.676 (0.000)	0.874 (0.000)	0.520 (0.000)	0.676 (0.000)	0.882 (0.000)	0.543 (0.000)	0.686 (0.000)

Table 4. Summary Statistics for df = bank-marketing

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.29	naive	0.579 (0.006)	**0.293 (0.003)**	0.150 (0.005)	0.579 (0.006)	**0.293 (0.003)**	0.150 (0.005)	0.579 (0.006)	**0.293 (0.003)**	0.150 (0.005)
0.29	clust	0.711 (0.009)	0.289 (0.005)	0.081 (0.004)	0.711 (0.009)	0.289 (0.005)	0.081 (0.004)	0.711 (0.009)	0.289 (0.005)	0.081 (0.004)
0.29	strict-lassclust	**0.720 (0.004)**	0.289 (0.002)	0.082 (0.004)	**0.720 (0.004)**	0.289 (0.002)	0.082 (0.004)	**0.720 (0.004)**	0.289 (0.002)	0.082 (0.004)
0.29	non-strict-lassclust	**0.720 (0.004)**	0.289 (0.002)	0.082 (0.004)	**0.720 (0.004)**	0.289 (0.002)	0.082 (0.004)	**0.720 (0.004)**	0.289 (0.002)	0.082 (0.004)
0.29	lassojoint	0.628 (0.040)	0.205 (0.096)	**0.154 (0.056)**	0.628 (0.040)	0.205 (0.096)	**0.154 (0.056)**	0.628 (0.040)	0.205 (0.096)	**0.154 (0.056)**
0.5	naive	0.586 (0.007)	0.368 (0.003)	**0.131 (0.002)**	0.586 (0.007)	0.368 (0.003)	**0.131 (0.002)**	0.586 (0.007)	0.368 (0.003)	**0.131 (0.002)**
0.5	clust	0.476 (0.134)	0.512 (0.022)	0.099 (0.005)	0.537 (0.150)	0.485 (0.069)	0.095 (0.009)	0.474 (0.136)	0.512 (0.029)	0.097 (0.010)
0.5	strict-lassclust	**0.677 (0.004)**	0.517 (0.002)	0.103 (0.002)	**0.677 (0.004)**	0.517 (0.002)	0.103 (0.002)	**0.677 (0.004)**	0.517 (0.002)	0.103 (0.002)
0.5	non-strict-lassclust	**0.677 (0.004)**	0.517 (0.002)	0.103 (0.002)	**0.677 (0.004)**	0.517 (0.002)	0.103 (0.002)	**0.677 (0.004)**	0.517 (0.002)	0.103 (0.002)
0.5	lassojoint	0.632 (0.004)	**0.531 (0.002)**	0.129 (0.003)	0.632 (0.004)	**0.531 (0.002)**	0.129 (0.003)	0.632 (0.004)	**0.531 (0.002)**	0.129 (0.003)
0.8	naive	0.615 (0.001)	**0.776 (0.001)**	**0.263 (0.006)**	0.615 (0.001)	**0.776 (0.001)**	**0.263 (0.006)**	0.615 (0.001)	**0.776 (0.001)**	**0.263 (0.006)**
0.8	clust	**0.659 (0.002)**	0.768 (0.002)	0.141 (0.003)	**0.659 (0.002)**	0.768 (0.002)	0.141 (0.003)	**0.659 (0.002)**	0.768 (0.002)	0.141 (0.003)
0.8	strict-lassclust	0.656 (0.002)	0.768 (0.003)	0.139 (0.003)	0.656 (0.002)	0.768 (0.003)	0.139 (0.003)	0.656 (0.002)	0.768 (0.003)	0.139 (0.003)
0.8	non-strict-lassclust	0.656 (0.002)	0.768 (0.003)	0.139 (0.003)	0.656 (0.002)	0.768 (0.003)	0.139 (0.003)	0.656 (0.002)	0.768 (0.003)	0.139 (0.003)
0.8	lassojoint	0.460 (0.003)	0.774 (0.001)	0.151 (0.009)	0.460 (0.003)	0.774 (0.001)	0.151 (0.009)	0.460 (0.003)	0.774 (0.001)	0.151 (0.009)

Table 5. Summary Statistics for df = banknote

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.29	naive	**0.738 (0.024)**	**0.572 (0.020)**	**0.671 (0.020)**	**0.738 (0.024)**	**0.572 (0.020)**	**0.671 (0.020)**	**0.738 (0.024)**	**0.572 (0.020)**	**0.671 (0.020)**
0.29	clust	0.697 (0.025)	0.398 (0.021)	0.425 (0.032)	0.632 (0.041)	0.400 (0.019)	0.392 (0.036)	0.575 (0.028)	0.425 (0.018)	0.386 (0.026)
0.29	strict-lassclust	0.704 (0.023)	0.397 (0.021)	0.428 (0.031)	0.658 (0.033)	0.400 (0.020)	0.397 (0.038)	0.603 (0.020)	0.425 (0.018)	0.386 (0.026)
0.29	non-strict-lassclust	0.703 (0.022)	0.398 (0.021)	0.423 (0.030)	0.658 (0.033)	0.400 (0.020)	0.396 (0.036)	0.603 (0.020)	0.425 (0.018)	0.386 (0.026)
0.29	lassojoint	0.620 (0.084)	0.500 (0.045)	0.569 (0.065)	0.620 (0.084)	0.500 (0.045)	0.569 (0.065)	0.622 (0.085)	0.501 (0.045)	0.570 (0.066)
0.45	naive	0.779 (0.004)	**0.650 (0.011)**	**0.703 (0.011)**	0.779 (0.004)	**0.650 (0.011)**	**0.703 (0.011)**	0.779 (0.004)	**0.650 (0.011)**	**0.703 (0.011)**
0.45	clust	0.592 (0.003)	0.448 (0.010)	0.418 (0.017)	0.617 (0.010)	0.441 (0.006)	0.403 (0.011)	0.614 (0.007)	0.427 (0.010)	0.368 (0.018)
0.45	strict-lassclust	0.593 (0.003)	0.448 (0.011)	0.419 (0.020)	0.619 (0.017)	0.439 (0.008)	0.405 (0.011)	0.633 (0.009)	0.427 (0.010)	0.368 (0.018)
0.45	non-strict-lassclust	0.592 (0.004)	0.447 (0.010)	0.418 (0.015)	0.616 (0.008)	0.442 (0.006)	0.404 (0.010)	0.633 (0.009)	0.427 (0.010)	0.368 (0.018)
0.45	lassojoint	**0.792 (0.010)**	0.616 (0.037)	0.691 (0.017)	**0.792 (0.010)**	0.616 (0.037)	0.691 (0.017)	**0.792 (0.010)**	0.616 (0.037)	0.691 (0.017)
0.92	naive	**0.991 (0.000)**	**0.952 (0.003)**	**0.947 (0.002)**	**0.991 (0.000)**	**0.952 (0.003)**	**0.947 (0.002)**	**0.991 (0.000)**	**0.952 (0.003)**	**0.947 (0.002)**
0.92	clust	0.767 (0.038)	0.743 (0.015)	0.567 (0.033)	0.776 (0.025)	0.767 (0.014)	0.622 (0.029)	0.838 (0.010)	0.804 (0.011)	0.702 (0.020)
0.92	strict-lassclust	0.777 (0.038)	0.749 (0.011)	0.581 (0.025)	0.787 (0.022)	0.765 (0.015)	0.618 (0.031)	0.840 (0.011)	0.804 (0.009)	0.700 (0.017)
0.92	non-strict-lassclust	0.780 (0.037)	0.742 (0.014)	0.566 (0.031)	0.787 (0.022)	0.765 (0.015)	0.618 (0.031)	0.840 (0.011)	0.804 (0.009)	0.700 (0.017)
0.92	lassojoint	0.948 (0.001)	0.822 (0.001)	0.825 (0.000)	0.948 (0.001)	0.822 (0.001)	0.825 (0.000)	0.948 (0.001)	0.822 (0.001)	0.825 (0.000)

Table 6. Summary Statistics for df = breastc

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.26	naive	0.939 (0.015)	0.456 (0.033)	0.559 (0.032)	0.939 (0.015)	0.456 (0.033)	0.559 (0.032)	0.939 (0.015)	0.456 (0.033)	0.559 (0.032)
0.26	clust	0.966 (0.014)	0.959 (0.013)	**0.940 (0.020)**	0.967 (0.014)	0.959 (0.013)	0.940 (0.019)	0.968 (0.014)	**0.959 (0.013)**	**0.940 (0.019)**
0.26	strict-lassclust	**0.994 (0.003)**	0.958 (0.014)	0.939 (0.020)	**0.994 (0.003)**	**0.959 (0.014)**	0.939 (0.020)	**0.994 (0.003)**	**0.959 (0.013)**	**0.940 (0.019)**
0.26	non-strict-lassclust	**0.994 (0.003)**	**0.959 (0.014)**	**0.940 (0.020)**	**0.994 (0.003)**	**0.959 (0.014)**	**0.940 (0.020)**	**0.994 (0.003)**	**0.959 (0.013)**	**0.940 (0.019)**
0.26	lassojoint	0.984 (0.011)	0.354 (0.033)	0.519 (0.032)	0.984 (0.011)	0.354 (0.033)	0.519 (0.032)	0.984 (0.011)	0.354 (0.033)	0.519 (0.032)
0.52	naive	0.958 (0.003)	0.772 (0.021)	0.736 (0.024)	0.958 (0.003)	0.772 (0.021)	0.736 (0.024)	0.958 (0.003)	0.772 (0.021)	0.736 (0.024)
0.52	clust	0.960 (0.005)	**0.961 (0.001)**	**0.937 (0.003)**	0.962 (0.004)	0.959 (0.003)	0.934 (0.002)	0.965 (0.011)	0.962 (0.005)	0.938 (0.010)
0.52	strict-lassclust	**0.994 (0.000)**	0.960 (0.003)	0.936 (0.006)	**0.993 (0.001)**	0.959 (0.003)	0.934 (0.004)	**0.992 (0.001)**	**0.965 (0.005)**	**0.944 (0.009)**
0.52	non-strict-lassclust	0.993 (0.001)	0.960 (0.003)	0.936 (0.006)	0.993 (0.000)	**0.961 (0.002)**	**0.937 (0.005)**	**0.992 (0.001)**	**0.965 (0.005)**	**0.944 (0.009)**
0.52	lassojoint	0.985 (0.002)	0.317 (0.011)	0.482 (0.013)	0.985 (0.003)	0.317 (0.011)	0.482 (0.013)	0.985 (0.002)	0.317 (0.011)	0.482 (0.013)
0.89	naive	0.981 (0.000)	0.942 (0.007)	0.920 (0.010)	0.981 (0.000)	0.942 (0.007)	0.920 (0.010)	0.981 (0.000)	0.942 (0.007)	0.920 (0.010)
0.89	clust	0.953 (0.007)	0.951 (0.000)	0.922 (0.000)	0.964 (0.002)	**0.970 (0.004)**	**0.953 (0.007)**	0.972 (0.002)	**0.973 (0.002)**	**0.958 (0.003)**
0.89	strict-lassclust	**0.997 (0.000)**	**0.953 (0.002)**	**0.925 (0.004)**	**0.997 (0.000)**	0.968 (0.002)	0.950 (0.004)	**0.997 (0.001)**	**0.973 (0.002)**	**0.958 (0.003)**
0.89	non-strict-lassclust	**0.997 (0.000)**	**0.953 (0.002)**	**0.925 (0.004)**	**0.997 (0.000)**	**0.970 (0.004)**	**0.953 (0.007)**	**0.997 (0.001)**	**0.973 (0.002)**	**0.958 (0.003)**
0.89	lassojoint	0.996 (0.001)	0.849 (0.005)	0.814 (0.003)	0.996 (0.001)	0.849 (0.005)	0.814 (0.003)	0.996 (0.001)	0.849 (0.005)	0.814 (0.003)

Table 7. Summary Statistics for df = credit-a

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.31	naive	**0.686 (0.034)**	0.552 (0.028)	**0.653 (0.028)**	**0.686 (0.034)**	0.552 (0.028)	**0.653 (0.028)**	**0.686 (0.034)**	0.552 (0.028)	**0.653 (0.028)**
0.31	clust	0.585 (0.028)	0.607 (0.028)	0.355 (0.064)	0.586 (0.024)	**0.609 (0.028)**	0.377 (0.064)	0.594 (0.023)	0.615 (0.028)	0.417 (0.050)
0.31	strict-lassclust	0.637 (0.034)	0.606 (0.028)	0.351 (0.066)	0.631 (0.033)	0.609 (0.027)	0.375 (0.063)	0.632 (0.030)	**0.616 (0.027)**	0.418 (0.050)
0.31	non-strict-lassclust	0.639 (0.035)	**0.608 (0.029)**	0.355 (0.068)	0.632 (0.034)	0.609 (0.027)	0.376 (0.062)	0.632 (0.030)	**0.616 (0.027)**	0.418 (0.050)
0.31	lassojoint	0.657 (0.034)	0.507 (0.044)	0.636 (0.030)	0.658 (0.035)	0.507 (0.045)	0.636 (0.031)	0.657 (0.034)	0.506 (0.044)	0.636 (0.030)
0.51	naive	**0.706 (0.016)**	0.584 (0.013)	**0.631 (0.016)**	**0.706 (0.016)**	0.584 (0.013)	**0.631 (0.016)**	**0.706 (0.016)**	0.584 (0.013)	**0.631 (0.016)**
0.51	clust	0.590 (0.009)	0.635 (0.014)	0.361 (0.013)	0.600 (0.015)	**0.647 (0.012)**	0.408 (0.039)	0.605 (0.013)	0.637 (0.010)	0.437 (0.029)
0.51	strict-lassclust	0.626 (0.011)	**0.639 (0.014)**	0.368 (0.012)	0.638 (0.021)	0.646 (0.009)	0.406 (0.043)	0.638 (0.020)	**0.639 (0.010)**	0.438 (0.028)
0.51	non-strict-lassclust	0.626 (0.012)	**0.639 (0.014)**	0.368 (0.012)	0.638 (0.021)	0.646 (0.009)	0.406 (0.043)	0.638 (0.020)	**0.639 (0.010)**	0.438 (0.028)
0.51	lassojoint	0.672 (0.022)	0.544 (0.060)	0.594 (0.022)	0.672 (0.022)	0.544 (0.060)	0.594 (0.022)	0.672 (0.022)	0.544 (0.060)	0.594 (0.022)
0.81	naive	**0.804 (0.000)**	0.718 (0.000)	**0.686 (0.000)**	**0.804 (0.000)**	0.718 (0.000)	**0.686 (0.000)**	**0.804 (0.000)**	0.718 (0.000)	**0.686 (0.000)**
0.81	clust	0.595 (0.000)	0.641 (0.000)	0.407 (0.000)	0.670 (0.000)	0.636 (0.000)	0.393 (0.000)	0.666 (0.000)	0.631 (0.000)	0.400 (0.000)
0.81	strict-lassclust	0.675 (0.000)	0.626 (0.000)	0.342 (0.000)	0.676 (0.000)	0.636 (0.000)	0.393 (0.000)	0.669 (0.000)	0.631 (0.000)	0.400 (0.000)
0.81	non-strict-lassclust	0.677 (0.000)	0.626 (0.000)	0.342 (0.000)	0.676 (0.000)	0.636 (0.000)	0.393 (0.000)	0.669 (0.000)	0.631 (0.000)	0.400 (0.000)
0.81	lassojoint	0.794 (0.000)	**0.723 (0.000)**	0.675 (0.000)	0.794 (0.000)	**0.723 (0.000)**	0.675 (0.000)	0.794 (0.000)	**0.723 (0.000)**	0.675 (0.000)

Table 8. Summary Statistics for df = credit-g

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.31	naive	0.524 (0.022)	0.340 (0.030)	**0.471 (0.027)**	0.524 (0.022)	0.340 (0.030)	**0.471 (0.027)**	0.524 (0.022)	0.340 (0.030)	**0.471 (0.027)**
0.31	clust	0.408 (0.022)	0.511 (0.025)	0.161 (0.032)	0.408 (0.023)	0.510 (0.025)	0.162 (0.032)	0.408 (0.023)	0.509 (0.025)	0.165 (0.033)
0.31	strict-lassclust	0.561 (0.112)	0.511 (0.025)	0.162 (0.031)	0.579 (0.104)	**0.511 (0.025)**	0.163 (0.032)	**0.571 (0.108)**	**0.510 (0.025)**	0.165 (0.033)
0.31	non-strict-lassclust	**0.565 (0.110)**	**0.512 (0.025)**	0.162 (0.031)	**0.581 (0.105)**	**0.511 (0.025)**	0.163 (0.032)	**0.571 (0.108)**	**0.510 (0.025)**	0.165 (0.033)
0.31	lassojoint	0.523 (0.026)	0.456 (0.073)	0.365 (0.065)	0.523 (0.026)	0.456 (0.073)	0.365 (0.065)	0.523 (0.026)	0.456 (0.073)	0.365 (0.065)
0.49	naive	0.518 (0.008)	0.514 (0.011)	**0.444 (0.010)**	0.518 (0.008)	0.514 (0.011)	**0.444 (0.010)**	0.518 (0.008)	0.514 (0.011)	**0.444 (0.010)**
0.49	clust	0.427 (0.006)	**0.548 (0.015)**	0.185 (0.001)	0.427 (0.006)	0.536 (0.013)	0.181 (0.001)	0.427 (0.007)	**0.534 (0.017)**	0.181 (0.001)
0.49	strict-lassclust	0.493 (0.091)	0.545 (0.018)	0.184 (0.001)	0.504 (0.094)	**0.538 (0.015)**	0.182 (0.000)	0.491 (0.085)	**0.534 (0.017)**	0.181 (0.001)
0.49	non-strict-lassclust	0.410 (0.014)	0.547 (0.016)	0.185 (0.001)	0.504 (0.094)	**0.538 (0.015)**	0.182 (0.000)	0.491 (0.085)	**0.534 (0.017)**	0.181 (0.001)
0.49	lassojoint	**0.523 (0.014)**	0.446 (0.020)	0.439 (0.010)	**0.523 (0.014)**	0.446 (0.020)	0.439 (0.010)	**0.523 (0.014)**	0.446 (0.020)	0.439 (0.010)
0.81	naive	**0.591 (0.000)**	0.587 (0.000)	0.340 (0.000)	**0.591 (0.000)**	0.587 (0.000)	0.340 (0.000)	**0.591 (0.000)**	0.587 (0.000)	0.340 (0.000)
0.81	clust	0.520 (0.000)	**0.620 (0.000)**	0.174 (0.000)	0.516 (0.000)	**0.620 (0.000)**	0.174 (0.000)	0.500 (0.000)	**0.620 (0.000)**	0.174 (0.000)
0.81	strict-lassclust	0.557 (0.000)	**0.620 (0.000)**	0.174 (0.000)	0.537 (0.000)	**0.620 (0.000)**	0.174 (0.000)	0.557 (0.000)	**0.620 (0.000)**	0.174 (0.000)
0.81	non-strict-lassclust	0.527 (0.000)	**0.620 (0.000)**	0.174 (0.000)	0.537 (0.000)	**0.620 (0.000)**	0.174 (0.000)	0.557 (0.000)	**0.620 (0.000)**	0.174 (0.000)
0.81	lassojoint	0.589 (0.000)	0.537 (0.000)	**0.455 (0.000)**	0.590 (0.000)	0.540 (0.000)	**0.457 (0.000)**	0.590 (0.000)	0.540 (0.000)	**0.457 (0.000)**

Table 9. Summary Statistics for df = dhfr

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.24	naive	0.517 (0.040)	0.441 (0.047)	0.478 (0.063)	0.517 (0.040)	0.441 (0.047)	0.478 (0.063)	0.517 (0.040)	0.441 (0.047)	0.478 (0.063)
0.24	clust	0.561 (0.045)	0.534 (0.075)	0.409 (0.097)	0.569 (0.066)	0.567 (0.054)	0.442 (0.071)	0.546 (0.048)	0.523 (0.056)	0.433 (0.067)
0.24	strict-lassclust	0.585 (0.047)	0.533 (0.134)	0.427 (0.202)	**0.598 (0.060)**	0.524 (0.129)	**0.591 (0.162)**	**0.613 (0.044)**	**0.564 (0.053)**	0.484 (0.055)
0.24	non-strict-lassclust	**0.585 (0.086)**	**0.564 (0.083)**	0.347 (0.181)	0.590 (0.071)	**0.573 (0.100)**	0.459 (0.196)	**0.613 (0.044)**	**0.564 (0.053)**	0.484 (0.055)
0.24	lassojoint	0.551 (0.081)	0.458 (0.061)	**0.505 (0.058)**	0.544 (0.079)	0.455 (0.070)	0.520 (0.055)	0.532 (0.054)	0.487 (0.065)	**0.517 (0.058)**
0.5	naive	0.652 (0.000)	0.567 (0.000)	0.580 (0.000)	**0.652 (0.000)**	0.567 (0.000)	0.580 (0.000)	0.652 (0.000)	0.567 (0.000)	0.580 (0.000)
0.5	clust	0.734 (0.000)	0.742 (0.000)	0.615 (0.000)	0.593 (0.000)	0.598 (0.000)	0.466 (0.000)	0.593 (0.000)	0.598 (0.000)	0.466 (0.000)
0.5	strict-lassclust	0.548 (0.000)	0.577 (0.000)	0.196 (0.000)	0.571 (0.000)	0.515 (0.000)	0.277 (0.000)	0.588 (0.000)	0.495 (0.000)	0.246 (0.000)
0.5	non-strict-lassclust	**0.908 (0.000)**	**0.814 (0.000)**	**0.743 (0.000)**	0.564 (0.000)	0.515 (0.000)	0.299 (0.000)	0.588 (0.000)	0.495 (0.000)	0.246 (0.000)
0.5	lassojoint	0.653 (0.000)	0.629 (0.000)	0.640 (0.000)	0.647 (0.000)	**0.670 (0.000)**	**0.667 (0.000)**	**0.661 (0.000)**	**0.680 (0.000)**	**0.680 (0.000)**
0.81	naive	0.772 (0.000)	**0.763 (0.000)**	**0.729 (0.000)**	0.772 (0.000)	**0.763 (0.000)**	**0.729 (0.000)**	0.772 (0.000)	0.763 (0.000)	**0.729 (0.000)**
0.81	clust	0.808 (0.000)	0.732 (0.000)	0.594 (0.000)	0.665 (0.000)	0.691 (0.000)	0.559 (0.000)	0.665 (0.000)	0.691 (0.000)	0.559 (0.000)
0.81	strict-lassclust	0.608 (0.000)	0.619 (0.000)	0.213 (0.000)	0.740 (0.000)	**0.763 (0.000)**	0.596 (0.000)	0.734 (0.000)	**0.794 (0.000)**	0.667 (0.000)
0.81	non-strict-lassclust	**0.818 (0.000)**	**0.763 (0.000)**	0.646 (0.000)	0.735 (0.000)	**0.763 (0.000)**	0.610 (0.000)	0.734 (0.000)	**0.794 (0.000)**	0.667 (0.000)
0.81	lassojoint	0.810 (0.000)	0.691 (0.000)	0.694 (0.000)	**0.826 (0.000)**	0.660 (0.000)	0.680 (0.000)	**0.812 (0.000)**	0.660 (0.000)	0.673 (0.000)

Table 10. Summary Statistics for df = diabetes

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	0.682 (0.030)	0.482 (0.031)	**0.550 (0.034)**	0.682 (0.030)	0.482 (0.031)	**0.550 (0.034)**	0.682 (0.030)	0.482 (0.031)	**0.550 (0.034)**
0.3	clust	0.680 (0.030)	0.674 (0.025)	0.501 (0.040)	0.677 (0.031)	0.674 (0.025)	0.504 (0.041)	0.683 (0.029)	0.674 (0.024)	0.510 (0.038)
0.3	strict-lassclust	0.686 (0.028)	**0.674 (0.026)**	0.502 (0.041)	0.685 (0.028)	0.674 (0.026)	0.506 (0.041)	0.685 (0.027)	**0.674 (0.025)**	0.510 (0.040)
0.3	non-strict-lassclust	0.686 (0.028)	**0.674 (0.026)**	0.503 (0.041)	0.685 (0.028)	**0.675 (0.026)**	0.507 (0.041)	0.685 (0.027)	**0.674 (0.025)**	0.510 (0.040)
0.3	lassojoint	**0.703 (0.029)**	0.389 (0.032)	0.528 (0.031)	**0.703 (0.029)**	0.389 (0.032)	0.528 (0.031)	**0.703 (0.029)**	0.389 (0.032)	0.528 (0.031)
0.52	naive	0.700 (0.006)	0.600 (0.001)	0.626 (0.003)	0.700 (0.006)	0.600 (0.001)	0.626 (0.003)	0.700 (0.006)	0.600 (0.001)	0.626 (0.003)
0.52	clust	0.669 (0.007)	**0.636 (0.005)**	0.482 (0.009)	0.674 (0.006)	0.636 (0.005)	0.489 (0.007)	0.678 (0.005)	**0.636 (0.005)**	0.507 (0.006)
0.52	strict-lassclust	0.667 (0.006)	0.627 (0.006)	0.463 (0.010)	0.669 (0.007)	**0.640 (0.005)**	0.498 (0.006)	0.675 (0.005)	**0.636 (0.005)**	0.501 (0.006)
0.52	non-strict-lassclust	0.667 (0.006)	0.631 (0.006)	0.473 (0.008)	0.668 (0.007)	**0.640 (0.005)**	0.492 (0.006)	0.675 (0.005)	**0.636 (0.005)**	0.501 (0.006)
0.52	lassojoint	**0.720 (0.005)**	0.602 (0.025)	**0.653 (0.013)**	**0.721 (0.005)**	0.594 (0.024)	**0.645 (0.012)**	**0.720 (0.005)**	0.602 (0.025)	**0.653 (0.013)**
0.79	naive	0.787 (0.000)	0.691 (0.000)	0.628 (0.000)	0.787 (0.000)	0.691 (0.000)	0.628 (0.000)	0.787 (0.000)	0.691 (0.000)	0.628 (0.000)
0.79	clust	0.688 (0.000)	**0.696 (0.000)**	0.539 (0.000)	0.694 (0.000)	**0.700 (0.000)**	0.543 (0.000)	0.704 (0.000)	**0.696 (0.000)**	0.539 (0.000)
0.79	strict-lassclust	0.671 (0.000)	0.683 (0.000)	0.490 (0.000)	0.691 (0.000)	0.691 (0.000)	0.523 (0.000)	0.699 (0.000)	**0.696 (0.000)**	0.539 (0.000)
0.79	non-strict-lassclust	0.680 (0.000)	**0.696 (0.000)**	0.539 (0.000)	0.695 (0.000)	0.696 (0.000)	0.545 (0.000)	0.699 (0.000)	**0.696 (0.000)**	0.539 (0.000)
0.79	lassojoint	**0.791 (0.000)**	0.691 (0.000)	**0.632 (0.000)**	**0.791 (0.000)**	0.691 (0.000)	**0.632 (0.000)**	**0.791 (0.000)**	0.691 (0.000)	**0.632 (0.000)**

Table 11. Summary Statistics for df = spambase

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.27	naive	0.774 (0.009)	0.473 (0.014)	0.596 (0.012)	0.774 (0.009)	0.473 (0.014)	0.596 (0.012)	0.774 (0.009)	0.473 (0.014)	0.596 (0.012)
0.27	clust	**0.858 (0.009)**	**0.804 (0.009)**	**0.737 (0.013)**	**0.859 (0.009)**	**0.804 (0.009)**	**0.739 (0.013)**	**0.858 (0.009)**	**0.805 (0.009)**	**0.744 (0.013)**
0.27	strict-lassclust	0.857 (0.009)	**0.804 (0.009)**	**0.737 (0.013)**	0.857 (0.009)	0.803 (0.009)	0.738 (0.013)	**0.858 (0.009)**	0.803 (0.009)	0.742 (0.012)
0.27	non-strict-lassclust	0.857 (0.009)	**0.804 (0.009)**	**0.737 (0.013)**	0.857 (0.009)	0.803 (0.009)	0.738 (0.013)	**0.858 (0.009)**	0.803 (0.009)	0.742 (0.012)
0.27	lassojoint	0.843 (0.013)	0.504 (0.139)	0.616 (0.068)	0.843 (0.013)	0.508 (0.141)	0.619 (0.070)	0.842 (0.013)	0.505 (0.140)	0.617 (0.068)
0.56	naive	0.785 (0.004)	0.721 (0.004)	0.731 (0.004)	0.785 (0.004)	0.721 (0.004)	0.731 (0.004)	0.785 (0.004)	0.721 (0.004)	0.731 (0.004)
0.56	clust	0.859 (0.002)	**0.808 (0.002)**	**0.735 (0.006)**	0.862 (0.004)	0.811 (0.002)	0.745 (0.004)	0.863 (0.004)	**0.811 (0.002)**	**0.753 (0.002)**
0.56	strict-lassclust	0.856 (0.004)	0.807 (0.003)	0.733 (0.007)	0.861 (0.004)	**0.812 (0.004)**	**0.747 (0.005)**	0.862 (0.003)	**0.811 (0.002)**	0.752 (0.002)
0.56	non-strict-lassclust	0.858 (0.002)	**0.808 (0.002)**	0.735 (0.005)	0.860 (0.004)	0.811 (0.003)	0.746 (0.005)	0.862 (0.003)	**0.811 (0.002)**	0.752 (0.002)
0.56	lassojoint	**0.872 (0.001)**	0.450 (0.003)	0.589 (0.002)	**0.872 (0.001)**	0.450 (0.003)	0.589 (0.002)	**0.872 (0.001)**	0.450 (0.003)	0.589 (0.002)
0.84	naive	**0.916 (0.002)**	**0.843 (0.001)**	**0.823 (0.000)**	**0.916 (0.002)**	**0.843 (0.001)**	**0.823 (0.000)**	**0.916 (0.002)**	**0.843 (0.001)**	**0.823 (0.000)**
0.84	clust	0.867 (0.008)	0.807 (0.008)	0.731 (0.012)	0.877 (0.005)	0.823 (0.001)	0.765 (0.001)	0.884 (0.003)	0.821 (0.003)	0.767 (0.003)
0.84	strict-lassclust	0.863 (0.008)	0.803 (0.007)	0.725 (0.013)	0.876 (0.005)	0.818 (0.002)	0.759 (0.003)	0.880 (0.003)	0.820 (0.001)	0.766 (0.001)
0.84	non-strict-lassclust	0.865 (0.007)	0.804 (0.007)	0.727 (0.011)	0.876 (0.005)	0.819 (0.002)	0.760 (0.002)	0.880 (0.003)	0.820 (0.001)	0.766 (0.001)
0.84	lassojoint	0.898 (0.000)	0.400 (0.001)	0.571 (0.001)	0.899 (0.000)	0.400 (0.001)	0.571 (0.001)	0.899 (0.001)	0.400 (0.001)	0.571 (0.001)

Table 12. Summary Statistics for df = wdbc

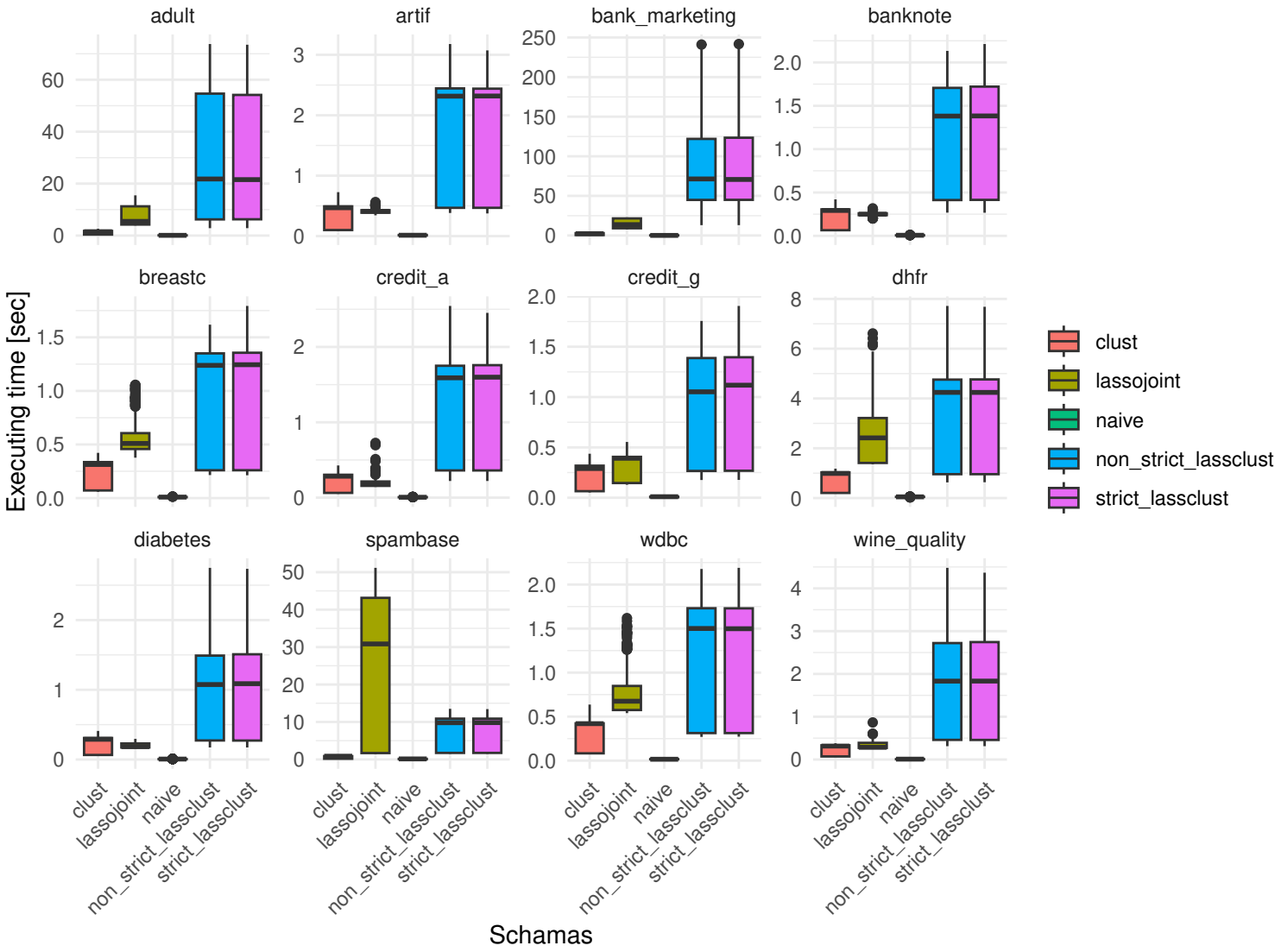
calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.24	naive	0.644 (0.027)	0.516 (0.035)	0.602 (0.031)	0.644 (0.027)	0.516 (0.035)	0.602 (0.031)	0.644 (0.027)	0.516 (0.035)	0.602 (0.031)
0.24	clust	0.918 (0.021)	0.905 (0.021)	0.862 (0.030)	0.920 (0.021)	0.905 (0.020)	0.863 (0.031)	0.922 (0.021)	0.908 (0.021)	0.869 (0.032)
0.24	strict-lassclust	0.967 (0.018)	0.853 (0.066)	0.739 (0.173)	0.965 (0.018)	0.866 (0.058)	0.775 (0.141)	0.967 (0.011)	**0.911 (0.019)**	**0.874 (0.028)**
0.24	non-strict-lassclust	0.970 (0.011)	**0.908 (0.029)**	**0.874 (0.034)**	0.968 (0.012)	**0.909 (0.026)**	**0.876 (0.033)**	0.967 (0.011)	**0.911 (0.019)**	**0.874 (0.028)**
0.24	lassojoint	**0.972 (0.013)**	0.394 (0.039)	0.552 (0.032)	**0.972 (0.013)**	0.394 (0.039)	0.552 (0.032)	**0.972 (0.013)**	0.398 (0.045)	0.553 (0.034)
0.52	naive	0.845 (0.000)	0.741 (0.000)	0.753 (0.000)	0.845 (0.000)	0.741 (0.000)	0.753 (0.000)	0.845 (0.000)	0.741 (0.000)	0.753 (0.000)
0.52	clust	0.948 (0.000)	0.900 (0.000)	0.857 (0.000)	0.941 (0.000)	0.900 (0.000)	0.860 (0.000)	0.944 (0.000)	**0.906 (0.000)**	**0.871 (0.000)**
0.52	strict-lassclust	**0.988 (0.000)**	0.894 (0.000)	0.845 (0.000)	0.983 (0.000)	0.706 (0.000)	0.405 (0.000)	0.983 (0.000)	0.900 (0.000)	0.862 (0.000)
0.52	non-strict-lassclust	0.984 (0.000)	**0.935 (0.000)**	**0.915 (0.000)**	0.982 (0.000)	**0.918 (0.000)**	**0.901 (0.000)**	0.983 (0.000)	0.900 (0.000)	0.862 (0.000)
0.52	lassojoint	0.987 (0.000)	0.535 (0.000)	0.629 (0.000)	**0.987 (0.000)**	0.535 (0.000)	0.629 (0.000)	**0.987 (0.000)**	0.535 (0.000)	0.629 (0.000)
0.89	naive	0.937 (0.000)	**0.918 (0.000)**	**0.889 (0.000)**	0.937 (0.000)	0.918 (0.000)	0.889 (0.000)	0.937 (0.000)	**0.918 (0.000)**	**0.889 (0.000)**
0.89	clust	0.920 (0.000)	0.912 (0.000)	0.851 (0.000)	0.946 (0.000)	0.918 (0.000)	0.870 (0.000)	0.931 (0.000)	0.912 (0.000)	0.862 (0.000)
0.89	strict-lassclust	0.936 (0.000)	0.671 (0.000)	0.067 (0.000)	0.969 (0.000)	0.741 (0.000)	0.389 (0.000)	**0.975 (0.000)**	0.912 (0.000)	0.860 (0.000)
0.89	non-strict-lassclust	0.966 (0.000)	**0.918 (0.000)**	0.883 (0.000)	**0.979 (0.000)**	**0.941 (0.000)**	**0.917 (0.000)**	**0.975 (0.000)**	0.912 (0.000)	0.860 (0.000)
0.89	lassojoint	**0.970 (0.000)**	0.371 (0.000)	0.520 (0.000)	0.969 (0.000)	0.371 (0.000)	0.520 (0.000)	0.969 (0.000)	0.371 (0.000)	0.520 (0.000)

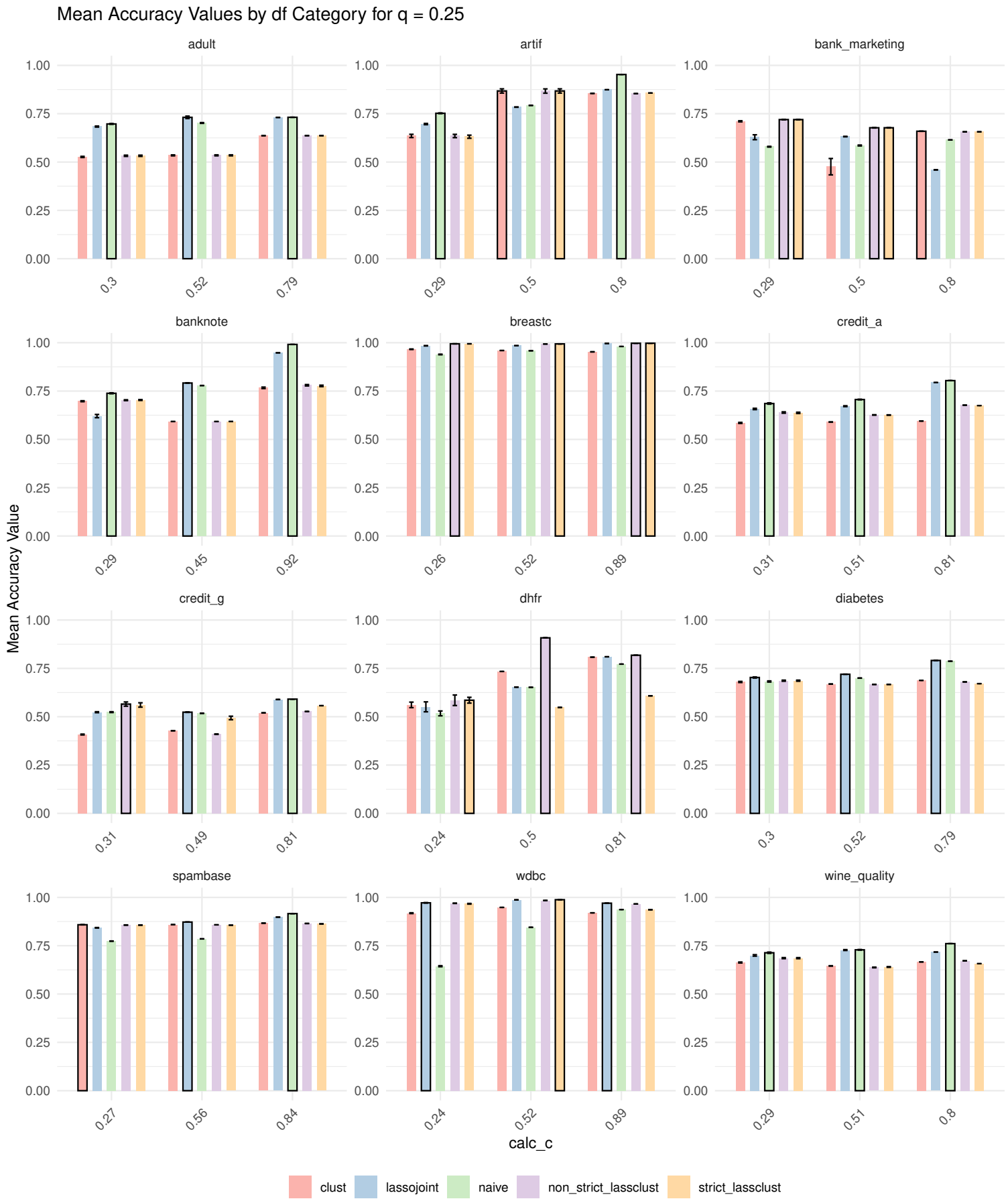
Table 13. Summary Statistics for df = wine-quality

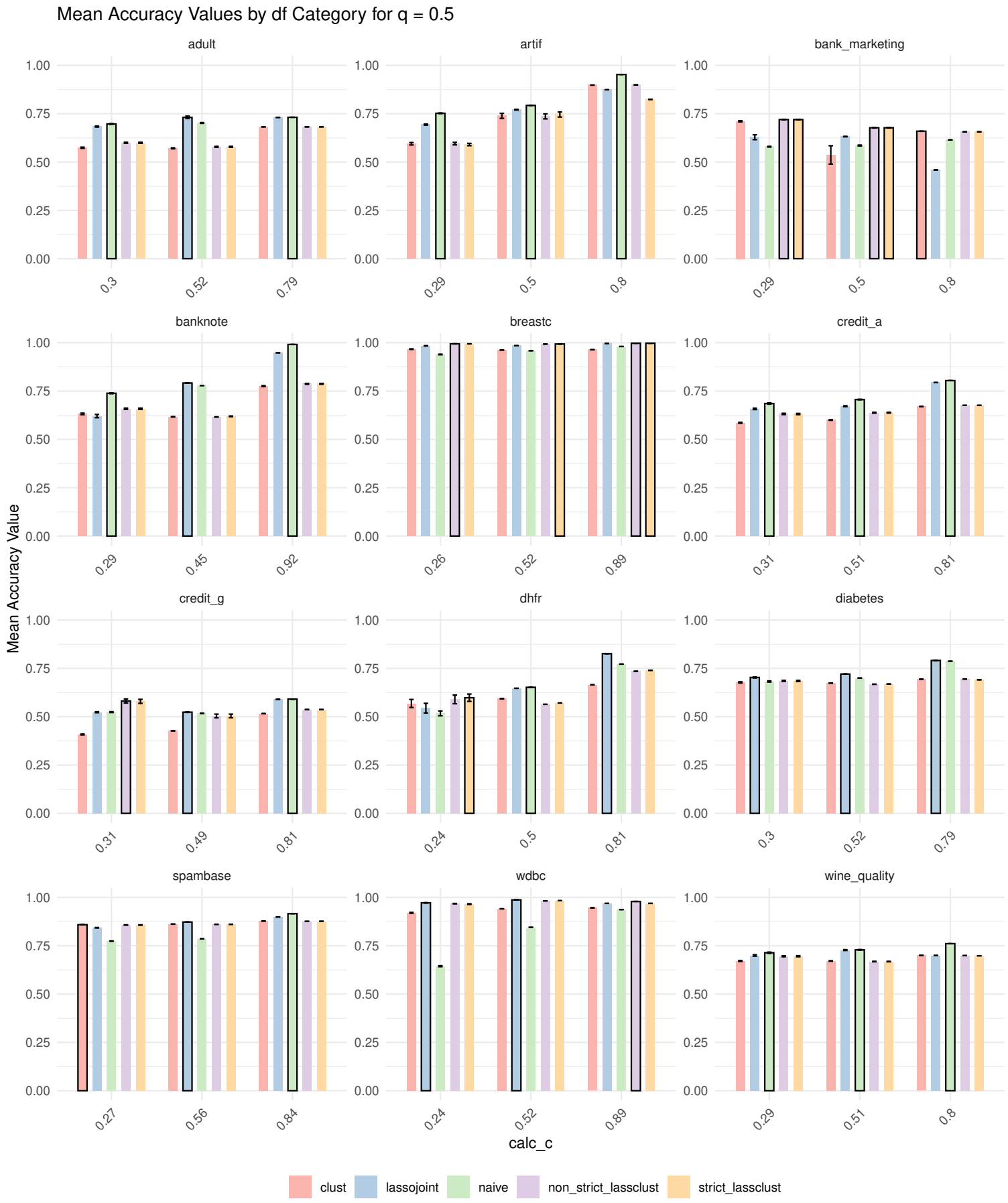
calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.29	naive	**0.714 (0.034)**	0.363 (0.019)	0.274 (0.023)	**0.714 (0.034)**	0.363 (0.019)	0.274 (0.023)	**0.714 (0.034)**	0.363 (0.019)	0.274 (0.023)
0.29	clust	0.663 (0.025)	0.704 (0.017)	**0.352 (0.030)**	0.671 (0.025)	**0.687 (0.019)**	**0.353 (0.030)**	0.678 (0.024)	**0.647 (0.019)**	**0.350 (0.029)**
0.29	strict-lassclust	0.686 (0.028)	**0.705 (0.017)**	**0.352 (0.030)**	0.696 (0.027)	**0.687 (0.019)**	**0.353 (0.030)**	0.691 (0.029)	**0.647 (0.019)**	**0.350 (0.029)**
0.29	non-strict-lassclust	0.686 (0.028)	0.704 (0.017)	0.352 (0.029)	0.695 (0.027)	**0.687 (0.019)**	**0.353 (0.030)**	0.691 (0.029)	**0.647 (0.019)**	**0.350 (0.029)**
0.29	lassojoint	0.700 (0.038)	0.261 (0.093)	0.255 (0.024)	0.700 (0.037)	0.251 (0.096)	0.254 (0.024)	0.701 (0.035)	0.252 (0.091)	0.254 (0.024)
0.51	naive	**0.729 (0.016)**	0.543 (0.007)	0.308 (0.011)	**0.729 (0.016)**	0.543 (0.007)	0.308 (0.011)	**0.729 (0.016)**	0.543 (0.007)	0.308 (0.011)
0.51	clust	0.645 (0.011)	**0.738 (0.008)**	0.312 (0.014)	0.671 (0.013)	**0.720 (0.019)**	**0.345 (0.023)**	0.676 (0.013)	0.679 (0.019)	0.366 (0.028)
0.51	strict-lassclust	0.640 (0.018)	0.735 (0.013)	0.310 (0.016)	0.668 (0.011)	0.720 (0.018)	**0.345 (0.023)**	0.678 (0.013)	**0.680 (0.018)**	**0.367 (0.027)**
0.51	non-strict-lassclust	0.638 (0.017)	0.734 (0.004)	**0.315 (0.014)**	0.668 (0.011)	**0.720 (0.019)**	**0.345 (0.023)**	0.678 (0.013)	**0.680 (0.018)**	**0.367 (0.027)**
0.51	lassojoint	0.728 (0.023)	0.517 (0.040)	0.298 (0.016)	0.728 (0.023)	0.517 (0.040)	0.298 (0.016)	0.713 (0.013)	0.527 (0.043)	0.302 (0.019)
0.8	naive	**0.761 (0.001)**	0.729 (0.002)	0.409 (0.004)	**0.761 (0.001)**	0.729 (0.002)	**0.409 (0.004)**	**0.761 (0.001)**	0.729 (0.002)	**0.409 (0.004)**
0.8	clust	0.666 (0.002)	0.749 (0.002)	0.326 (0.003)	0.700 (0.001)	0.756 (0.001)	0.400 (0.000)	0.711 (0.001)	0.745 (0.000)	0.396 (0.001)
0.8	strict-lassclust	0.657 (0.001)	0.754 (0.001)	0.314 (0.003)	0.698 (0.001)	0.756 (0.001)	0.394 (0.001)	0.711 (0.001)	0.745 (0.001)	0.396 (0.001)
0.8	non-strict-lassclust	0.672 (0.001)	0.745 (0.001)	0.344 (0.004)	0.699 (0.001)	0.754 (0.001)	0.392 (0.001)	0.711 (0.001)	0.745 (0.001)	0.396 (0.001)
0.8	lassojoint	0.717 (0.002)	**0.754 (0.006)**	**0.416 (0.003)**	0.700 (0.004)	**0.765 (0.007)**	0.398 (0.001)	0.700 (0.004)	**0.765 (0.007)**	0.398 (0.001)

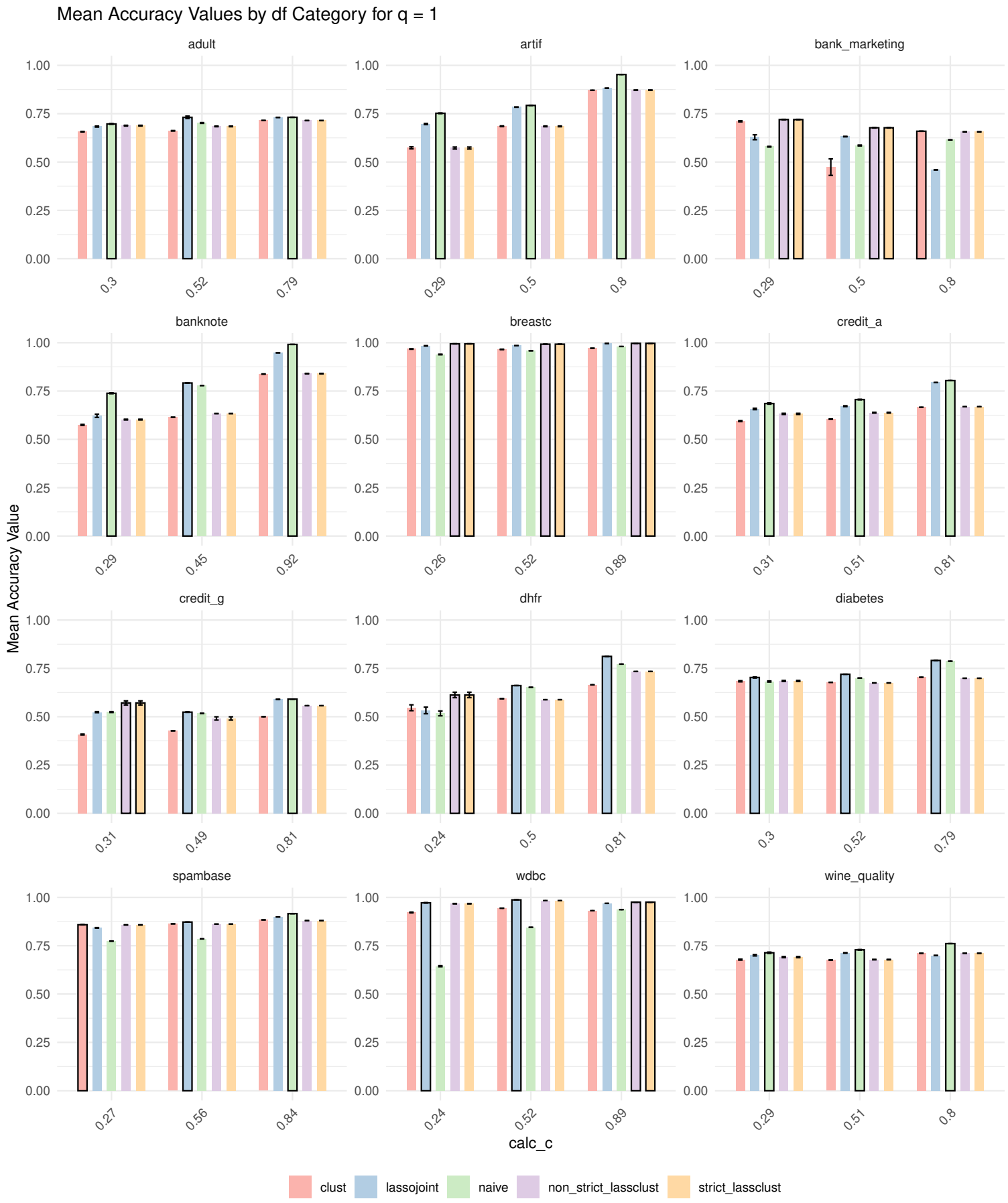
1.2 Graphs

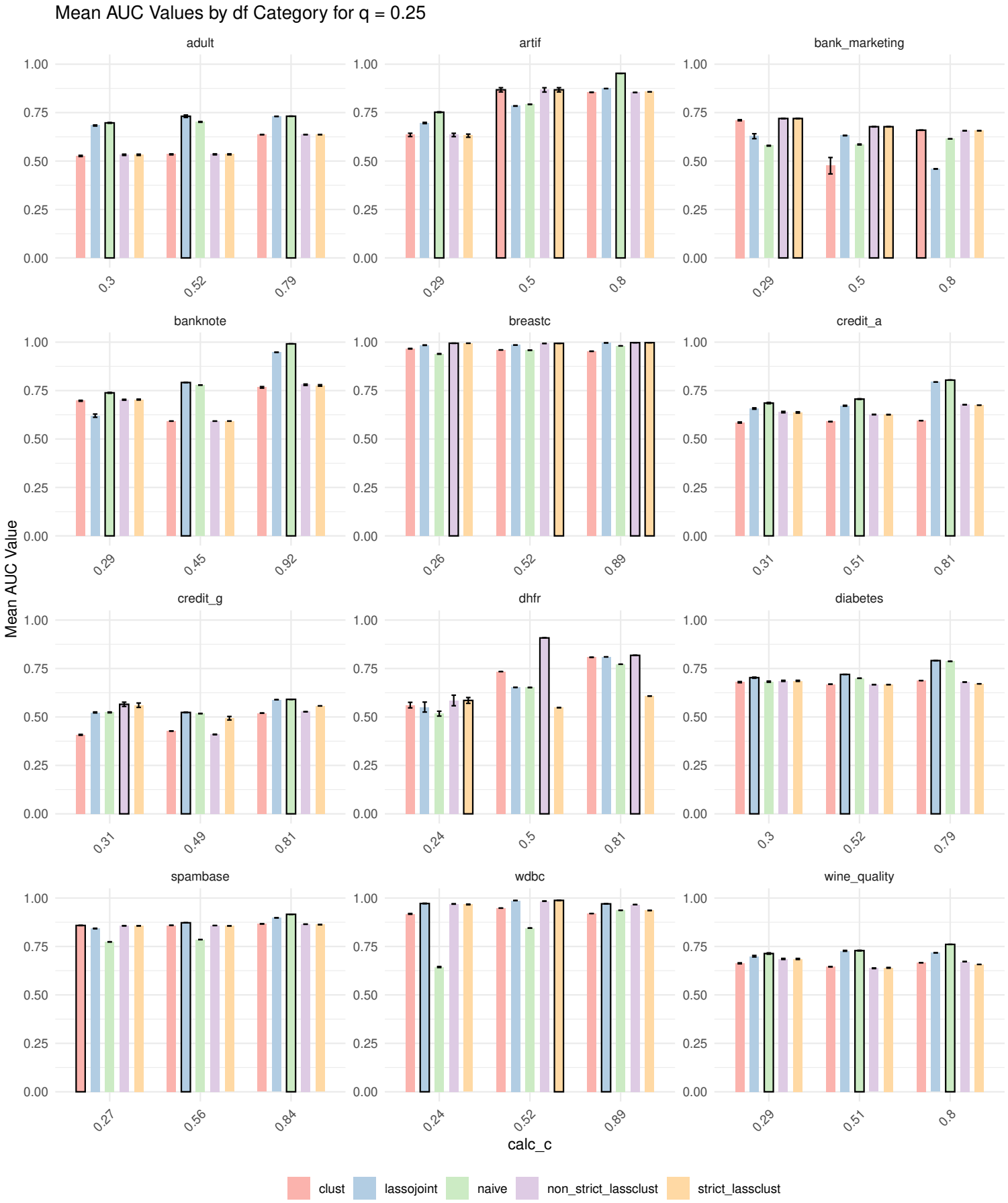
In this section we present all barplots for classification metrics for our methods. In addition, we present boxplots of the execution times of the algorithms used.

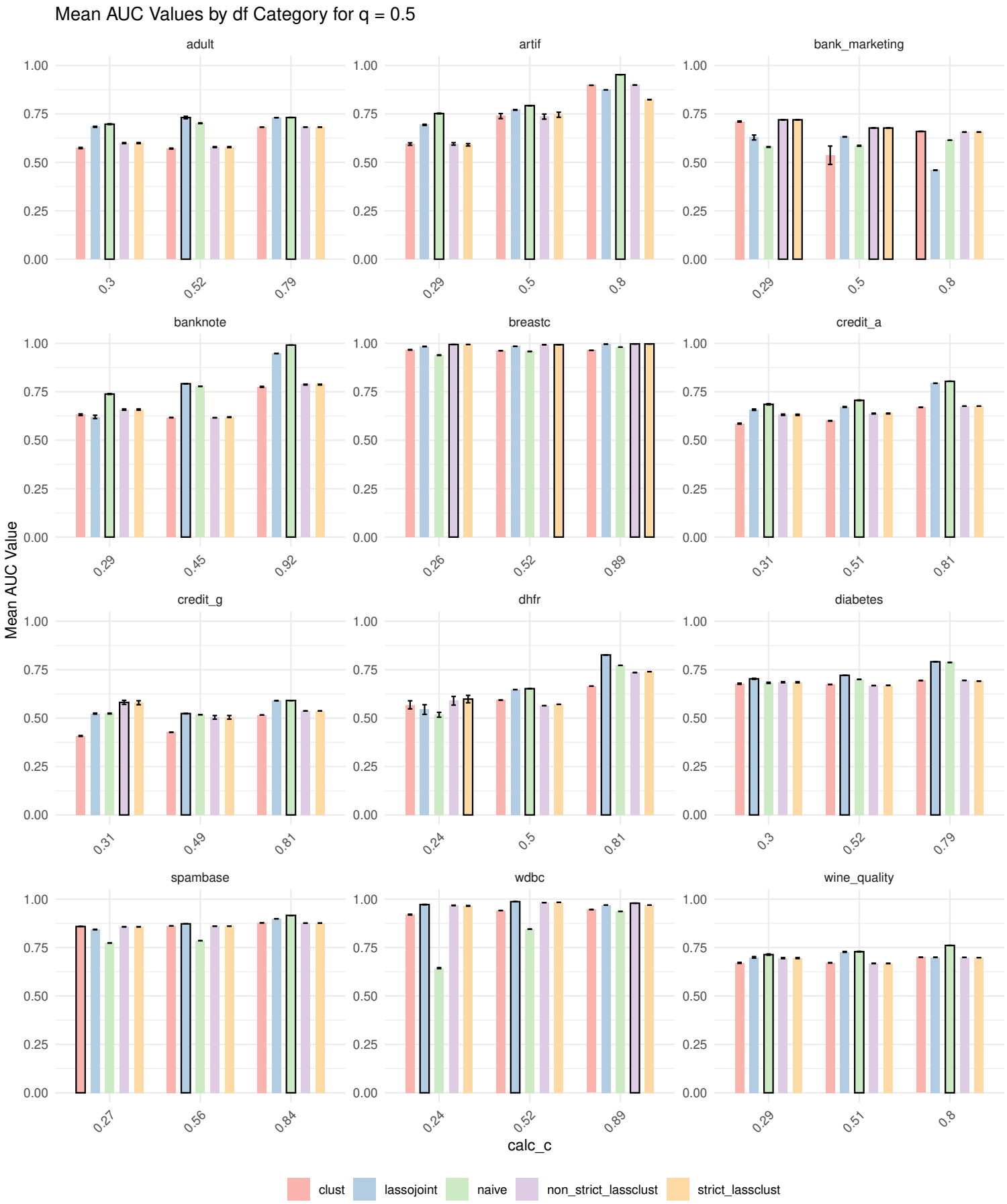

adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Executing time [sec]
Schamas
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust

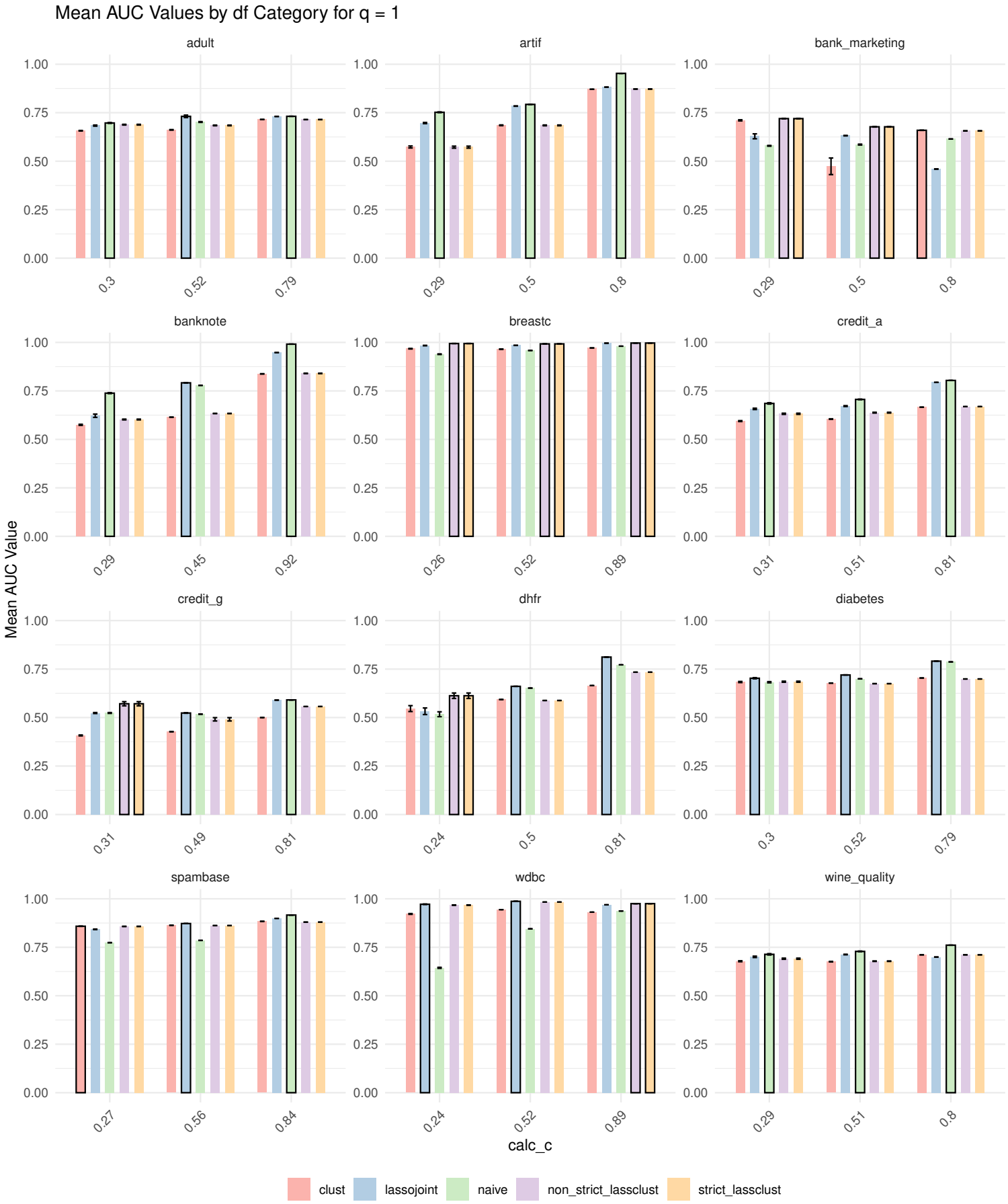

Mean Accuracy Values by df Category for q = 0.25
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean Accuracy Value
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust

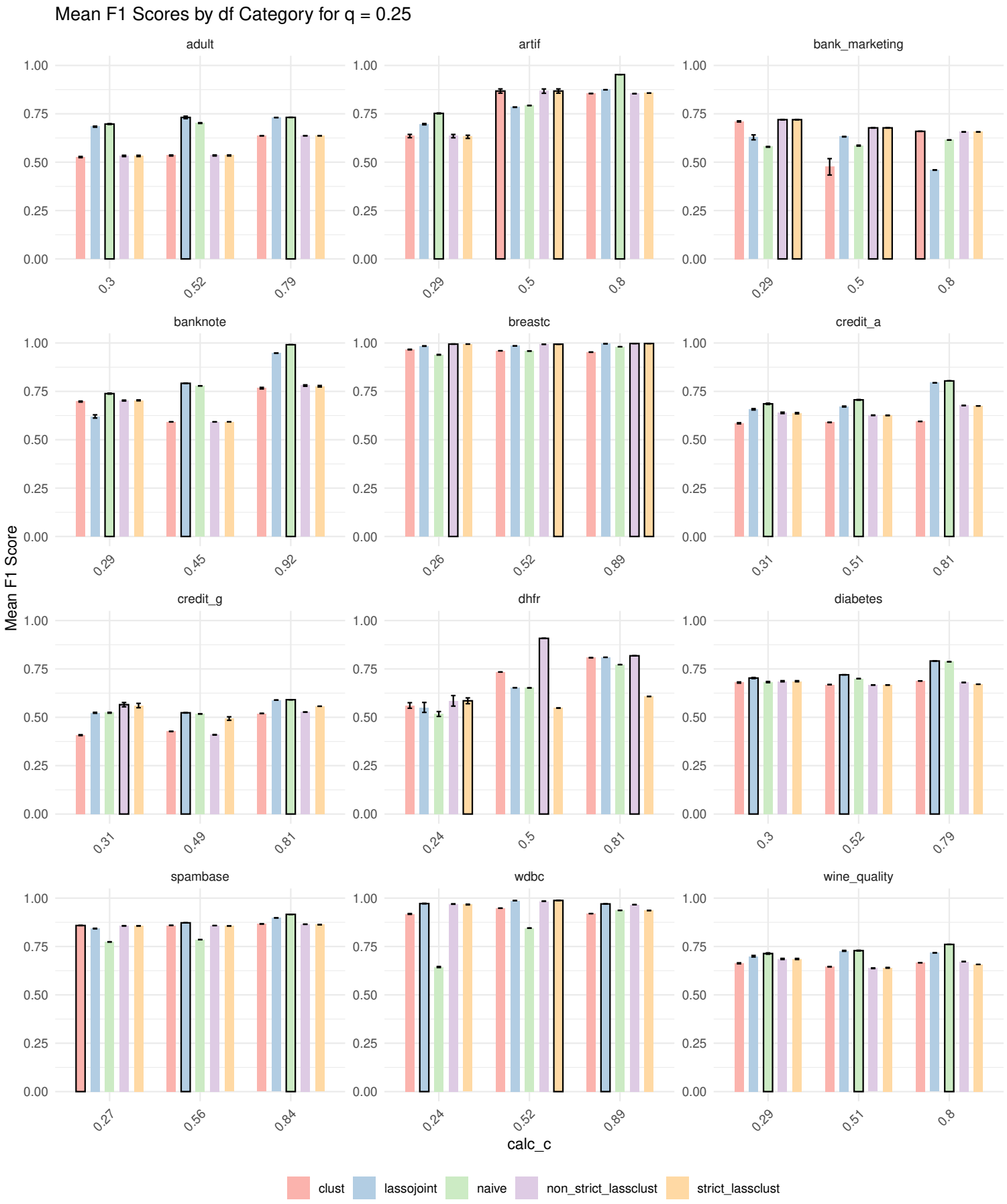

Mean Accuracy Values by df Category for q = 0.5
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean Accuracy Value
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust

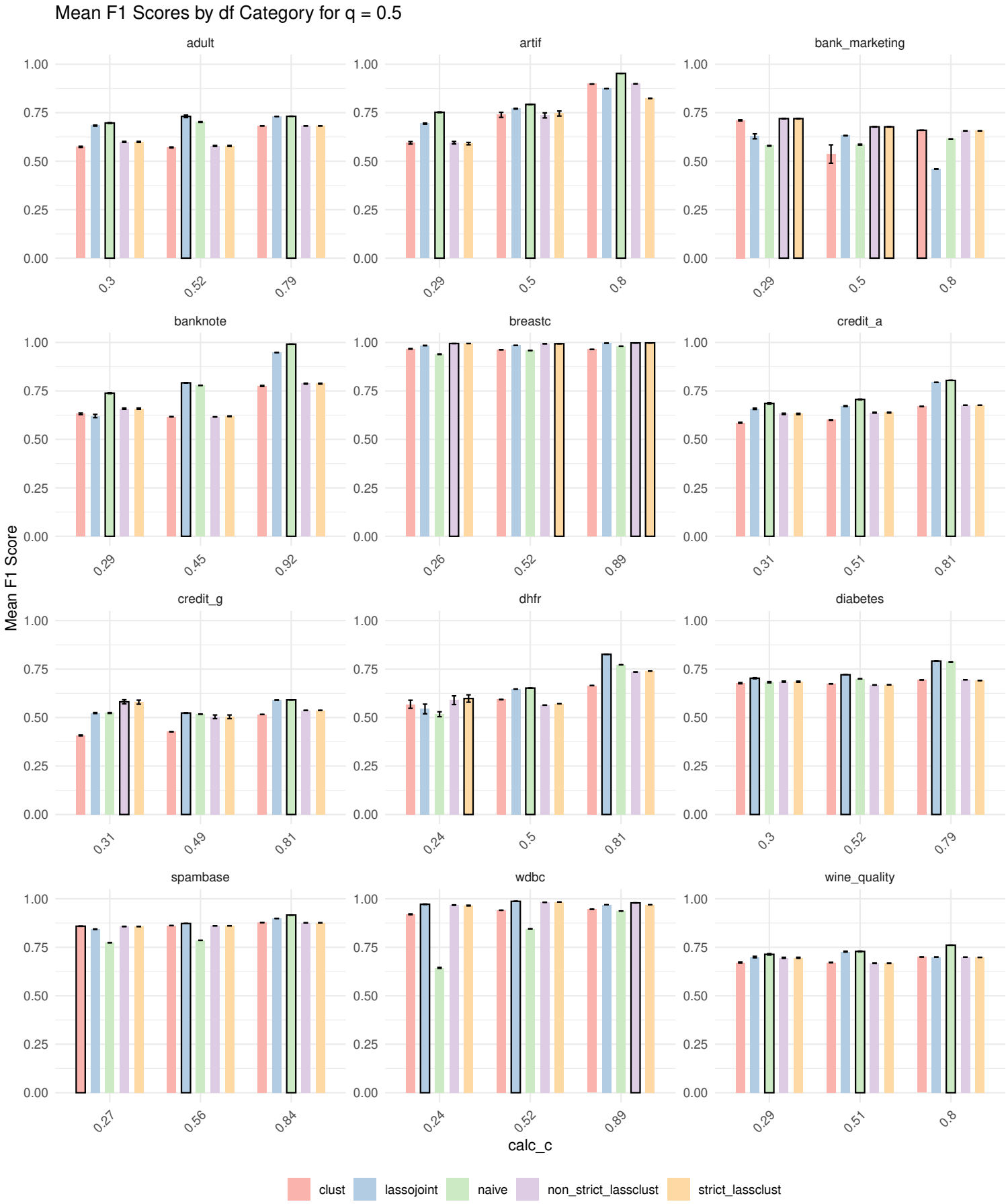

Mean Accuracy Values by df Category for q = 1
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean Accuracy Value
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust

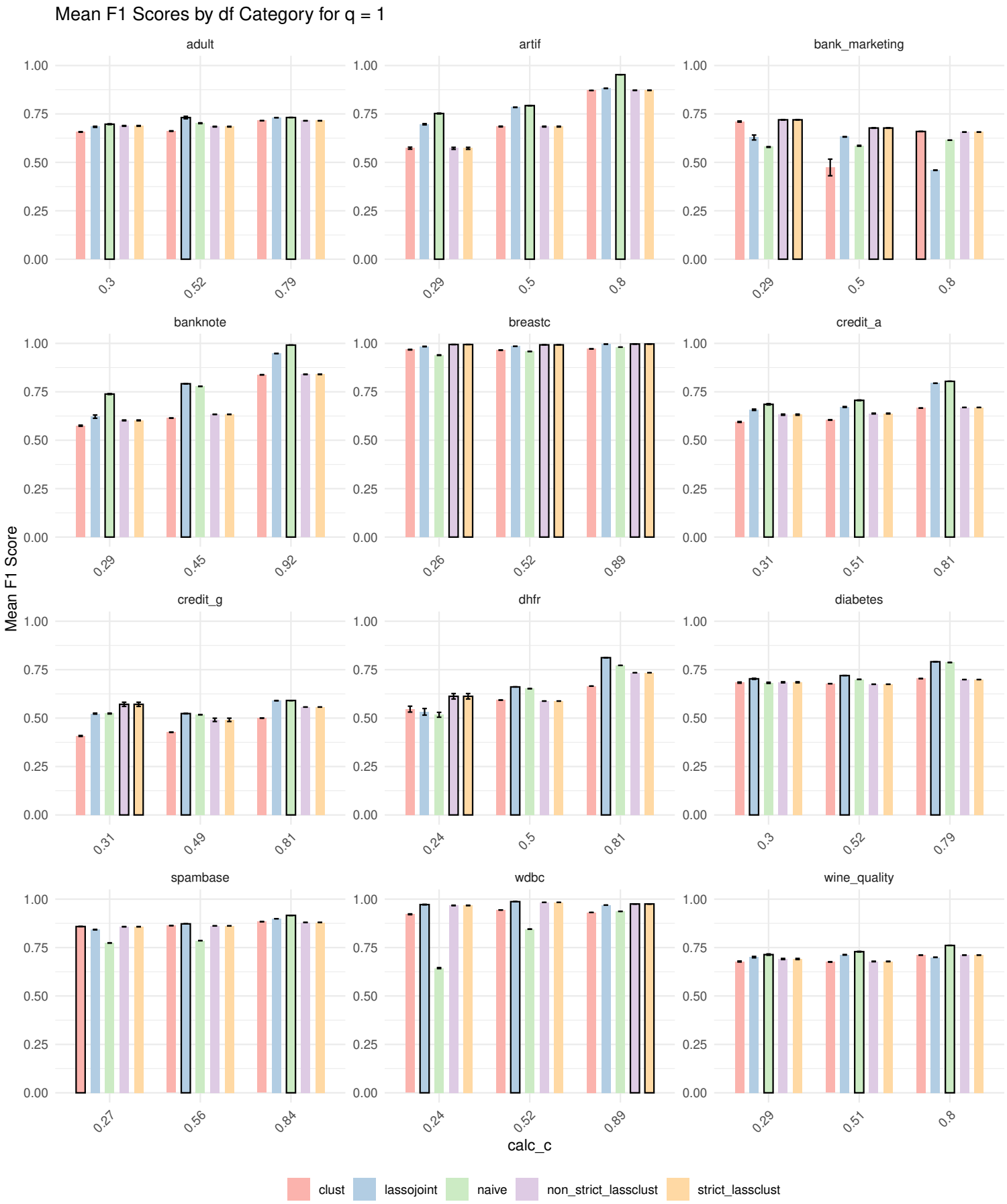

Mean AUC Values by df Category for q = 0.25
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean AUC Value
1.00
0.75
0.50
0.25
0.00
0.3
0.52
0.79
0.29
0.5
0.8
0.45
0.92
0.26
0.89
0.31
0.51
0.81
0.49
0.24
0.27
0.56
0.84
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust


Mean AUC Values by df Category for q = 0.5
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean AUC Value
1.00
0.75
0.50
0.25
0.00
0.3
0.52
0.79
0.29
0.5
0.8
0.45
0.92
0.26
0.89
0.31
0.51
0.81
0.49
0.24
0.27
0.56
0.84
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust


Mean AUC Values by df Category for q = 1
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean AUC Value
1.00
0.75
0.50
0.25
0.00
0.3
0.52
0.79
0.29
0.5
0.8
0.45
0.92
0.26
0.89
0.31
0.51
0.81
0.49
0.24
0.27
0.56
0.84
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust


Mean F1 Scores by df Category for q = 0.25
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean F1 Score
1.00
0.75
0.50
0.25
0.00
0.3
0.52
0.79
0.29
0.5
0.8
0.45
0.92
0.26
0.89
0.31
0.51
0.81
0.49
0.24
0.27
0.56
0.84
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust


Mean F1 Scores by df Category for q = 0.5
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean F1 Score
1.00
0.75
0.50
0.25
0.00
0.3
0.52
0.79
0.29
0.5
0.8
0.45
0.92
0.26
0.89
0.31
0.51
0.81
0.49
0.24
0.27
0.56
0.84
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust


Mean F1 Scores by df Category for q = 1
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean F1 Score
1.00
0.75
0.50
0.25
0.00
0.3
0.52
0.79
0.29
0.5
0.8
0.45
0.92
0.26
0.89
0.31
0.51
0.81
0.49
0.24
0.27
0.56
0.84
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust

2 SCAR Scheme

2.1 Tables for all classification metrics for 12 datasets.

q	naive acc	clust acc	strict lassclust acc	non strict lassclust acc	lassojoint acc	naive f1	clust f1	strict lassclust f1	non strict lassclust f1	lassojoint f1	naive auc	clust auc	strict lassclust auc	non strict lassclust auc	lassojoint auc
0.25	**0.58 (0.25)**	**0.76 (0.12)**	**0.75 (0.12)**	**0.76 (0.13)**	**0.79 (0.21)**	**0.61 (0.2)**	0.58 (0.21)	0.57 (0.21)	0.57 (0.21)	**0.69 (0.24)**	**0.9 (0.1)**	**0.81 (0.11)**	0.81 (0.12)	**0.82 (0.11)**	**0.9 (0.1)**
0.5	**0.58 (0.25)**	**0.76 (0.13)**	**0.75 (0.12)**	0.75 (0.13)	**0.79 (0.21)**	**0.61 (0.2)**	**0.59 (0.2)**	0.57 (0.2)	0.58 (0.2)	0.68 (0.25)	**0.9 (0.1)**	**0.81 (0.11)**	**0.82 (0.11)**	**0.82 (0.11)**	**0.9 (0.1)**
1	**0.58 (0.25)**	0.75 (0.13)	**0.75 (0.13)**	0.75 (0.13)	**0.79 (0.22)**	**0.61 (0.2)**	**0.59 (0.2)**	**0.59 (0.2)**	**0.59 (0.2)**	**0.69 (0.25)**	**0.9 (0.1)**	0.8 (0.11)	0.81 (0.12)	0.81 (0.12)	**0.9 (0.1)**

Table 14. Summary of Mean and Standard Deviation for Accuracy, F1, and AUC Metrics by q (Highest Mean in Bold); SCAR scheme

Table 15. Summary Statistics for df = adult; SCAR scheme

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	**0.797 (0.003)**	0.241 (0.005)	0.388 (0.006)	**0.797 (0.003)**	0.241 (0.005)	0.388 (0.006)	**0.797 (0.003)**	0.241 (0.005)	0.388 (0.006)
0.3	clust	0.740 (0.006)	0.697 (0.003)	0.477 (0.009)	0.738 (0.006)	0.695 (0.004)	0.473 (0.008)	0.734 (0.005)	0.693 (0.004)	0.469 (0.007)
0.3	strict-lassclust	0.740 (0.006)	0.697 (0.003)	0.475 (0.009)	0.737 (0.006)	0.695 (0.003)	0.472 (0.009)	0.733 (0.005)	0.693 (0.004)	0.469 (0.008)
0.3	non-strict-lassclust	0.740 (0.006)	0.697 (0.003)	0.476 (0.009)	0.737 (0.006)	0.695 (0.003)	0.472 (0.009)	0.733 (0.005)	0.693 (0.004)	0.469 (0.008)
0.3	lassojoint	0.795 (0.003)	**0.792 (0.003)**	**0.480 (0.033)**	0.795 (0.003)	**0.792 (0.003)**	**0.480 (0.033)**	0.795 (0.003)	**0.792 (0.003)**	**0.480 (0.033)**
0.5	naive	**0.799 (0.003)**	0.335 (0.005)	0.422 (0.003)	**0.799 (0.003)**	0.335 (0.005)	0.422 (0.003)	**0.799 (0.003)**	0.335 (0.005)	0.422 (0.003)
0.5	clust	0.767 (0.001)	0.732 (0.003)	0.483 (0.002)	0.762 (0.001)	0.727 (0.003)	0.471 (0.004)	0.756 (0.002)	0.721 (0.002)	0.459 (0.007)
0.5	strict-lassclust	0.766 (0.001)	0.731 (0.003)	0.482 (0.003)	0.761 (0.002)	0.727 (0.003)	0.470 (0.005)	0.756 (0.002)	0.721 (0.002)	0.457 (0.006)
0.5	non-strict-lassclust	0.766 (0.001)	0.732 (0.003)	0.482 (0.003)	0.761 (0.002)	0.727 (0.003)	0.470 (0.005)	0.756 (0.002)	0.721 (0.002)	0.457 (0.006)
0.5	lassojoint	0.797 (0.003)	**0.791 (0.001)**	**0.483 (0.020)**	0.797 (0.003)	**0.791 (0.001)**	**0.483 (0.020)**	0.797 (0.003)	**0.791 (0.001)**	**0.483 (0.020)**
0.8	naive	**0.797 (0.002)**	0.778 (0.003)	**0.532 (0.005)**	**0.797 (0.002)**	0.778 (0.003)	**0.532 (0.005)**	**0.797 (0.002)**	0.778 (0.003)	**0.532 (0.005)**
0.8	clust	0.790 (0.002)	0.775 (0.002)	0.434 (0.006)	0.785 (0.003)	0.768 (0.003)	0.409 (0.011)	0.777 (0.003)	0.760 (0.003)	0.373 (0.007)
0.8	strict-lassclust	0.790 (0.002)	0.774 (0.003)	0.432 (0.005)	0.784 (0.003)	0.768 (0.004)	0.408 (0.011)	0.777 (0.003)	0.760 (0.003)	0.369 (0.006)
0.8	non-strict-lassclust	0.790 (0.002)	0.774 (0.003)	0.431 (0.006)	0.784 (0.003)	0.768 (0.004)	0.408 (0.011)	0.777 (0.003)	0.760 (0.003)	0.369 (0.006)
0.8	lassojoint	0.796 (0.001)	**0.793 (0.003)**	0.462 (0.004)	0.796 (0.001)	**0.793 (0.003)**	0.462 (0.004)	0.796 (0.001)	**0.793 (0.003)**	0.462 (0.004)

Table 16. Summary Statistics for df = artif; SCAR scheme

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	0.968 (0.010)	0.520 (0.019)	0.670 (0.016)	0.968 (0.010)	0.520 (0.019)	0.670 (0.016)	0.968 (0.010)	0.520 (0.019)	0.670 (0.016)
0.3	clust	0.733 (0.082)	0.651 (0.051)	0.573 (0.059)	0.731 (0.077)	0.651 (0.045)	0.575 (0.048)	0.625 (0.052)	0.589 (0.039)	0.539 (0.047)
0.3	strict-lassclust	0.724 (0.085)	0.632 (0.060)	0.578 (0.111)	0.723 (0.077)	0.636 (0.051)	0.578 (0.102)	0.624 (0.052)	0.588 (0.040)	0.537 (0.048)
0.3	non-strict-lassclust	0.733 (0.082)	0.643 (0.047)	0.546 (0.077)	0.731 (0.077)	0.644 (0.044)	0.555 (0.060)	0.624 (0.052)	0.588 (0.040)	0.537 (0.048)
0.3	lassojoint	**0.989 (0.006)**	**0.944 (0.017)**	**0.942 (0.017)**	**0.989 (0.007)**	**0.944 (0.015)**	**0.943 (0.016)**	**0.988 (0.012)**	**0.943 (0.018)**	**0.941 (0.019)**
0.5	naive	0.988 (0.001)	0.671 (0.019)	0.754 (0.013)	0.988 (0.001)	0.671 (0.019)	0.754 (0.013)	0.988 (0.001)	0.671 (0.019)	0.754 (0.013)
0.5	clust	0.843 (0.086)	0.659 (0.012)	0.537 (0.054)	0.876 (0.046)	0.703 (0.010)	0.602 (0.009)	0.742 (0.017)	0.662 (0.011)	0.589 (0.014)
0.5	strict-lassclust	0.844 (0.088)	0.736 (0.104)	0.661 (0.187)	0.876 (0.046)	0.703 (0.014)	0.604 (0.019)	0.740 (0.018)	0.660 (0.011)	0.585 (0.013)
0.5	non-strict-lassclust	0.840 (0.086)	0.637 (0.036)	0.484 (0.125)	0.875 (0.046)	0.701 (0.012)	0.598 (0.011)	0.740 (0.018)	0.660 (0.011)	0.585 (0.013)
0.5	lassojoint	**0.995 (0.001)**	**0.956 (0.007)**	**0.956 (0.007)**	**0.995 (0.001)**	**0.960 (0.002)**	**0.960 (0.002)**	**0.995 (0.001)**	**0.960 (0.002)**	**0.960 (0.002)**
0.8	naive	**0.994 (0.003)**	**0.894 (0.026)**	**0.907 (0.022)**	0.994 (0.003)	**0.894 (0.026)**	**0.907 (0.022)**	0.994 (0.003)	**0.894 (0.026)**	**0.907 (0.022)**
0.8	clust	0.869 (0.050)	0.696 (0.023)	0.593 (0.031)	0.902 (0.012)	0.680 (0.016)	0.552 (0.034)	0.813 (0.009)	0.694 (0.006)	0.618 (0.008)
0.8	strict-lassclust	0.866 (0.046)	0.683 (0.052)	0.585 (0.143)	0.898 (0.018)	0.662 (0.026)	0.520 (0.065)	0.814 (0.009)	0.692 (0.007)	0.611 (0.007)
0.8	non-strict-lassclust	0.868 (0.050)	0.694 (0.040)	0.582 (0.077)	0.903 (0.013)	0.675 (0.013)	0.543 (0.028)	0.814 (0.009)	0.692 (0.007)	0.611 (0.007)
0.8	lassojoint	0.992 (0.004)	0.781 (0.175)	0.840 (0.119)	**0.995 (0.001)**	0.755 (0.199)	0.828 (0.131)	**0.995 (0.001)**	0.755 (0.199)	0.828 (0.131)

Table 17. Summary Statistics for df = bank-marketing; SCAR scheme

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	0.890 (0.023)	0.111 (0.006)	0.200 (0.010)	0.890 (0.023)	0.111 (0.006)	0.200 (0.010)	0.890 (0.023)	0.111 (0.006)	0.200 (0.010)
0.3	clust	0.827 (0.020)	0.747 (0.066)	0.357 (0.124)	0.828 (0.033)	0.721 (0.027)	0.360 (0.063)	0.831 (0.016)	0.718 (0.011)	0.355 (0.021)
0.3	strict-lassclust	0.831 (0.018)	0.747 (0.065)	0.369 (0.132)	0.833 (0.022)	0.715 (0.050)	0.353 (0.035)	0.835 (0.016)	0.718 (0.011)	0.356 (0.021)
0.3	non-strict-lassclust	0.823 (0.040)	0.744 (0.062)	0.337 (0.079)	0.833 (0.020)	0.721 (0.026)	0.362 (0.063)	0.835 (0.016)	0.718 (0.011)	0.356 (0.021)
0.3	lassojoint	**0.894 (0.032)**	**0.802 (0.239)**	**0.440 (0.109)**	**0.895 (0.032)**	**0.799 (0.241)**	**0.437 (0.110)**	**0.895 (0.025)**	**0.797 (0.244)**	**0.432 (0.113)**
0.5	naive	0.914 (0.005)	0.417 (0.033)	0.262 (0.008)	0.914 (0.005)	0.417 (0.033)	0.262 (0.008)	0.914 (0.005)	0.417 (0.033)	0.262 (0.008)
0.5	clust	0.841 (0.008)	0.753 (0.016)	0.358 (0.011)	0.841 (0.008)	0.753 (0.016)	0.358 (0.011)	0.841 (0.008)	0.753 (0.016)	0.358 (0.011)
0.5	strict-lassclust	0.843 (0.008)	0.753 (0.020)	0.360 (0.007)	0.843 (0.009)	0.751 (0.018)	0.359 (0.012)	0.843 (0.009)	0.751 (0.018)	0.359 (0.012)
0.5	non-strict-lassclust	0.843 (0.009)	0.753 (0.020)	0.361 (0.011)	0.843 (0.009)	0.753 (0.016)	0.361 (0.013)	0.843 (0.009)	0.751 (0.018)	0.359 (0.012)
0.5	lassojoint	**0.916 (0.004)**	**0.915 (0.003)**	**0.500 (0.030)**	**0.915 (0.004)**	**0.914 (0.003)**	**0.491 (0.027)**	**0.916 (0.004)**	**0.915 (0.003)**	**0.500 (0.032)**
0.8	naive	**0.917 (0.002)**	0.902 (0.001)	**0.509 (0.011)**	**0.917 (0.002)**	0.902 (0.001)	**0.509 (0.011)**	**0.917 (0.002)**	0.902 (0.001)	**0.509 (0.011)**
0.8	clust	0.862 (0.002)	0.900 (0.002)	0.483 (0.009)	0.862 (0.002)	0.900 (0.002)	0.483 (0.009)	0.862 (0.002)	0.900 (0.002)	0.483 (0.009)
0.8	strict-lassclust	0.863 (0.002)	0.902 (0.001)	0.490 (0.003)	0.863 (0.002)	0.901 (0.002)	0.485 (0.008)	0.863 (0.002)	0.902 (0.002)	0.485 (0.009)
0.8	non-strict-lassclust	0.863 (0.002)	0.902 (0.001)	0.490 (0.003)	0.863 (0.002)	0.901 (0.002)	0.485 (0.008)	0.863 (0.002)	0.902 (0.002)	0.485 (0.009)
0.8	lassojoint	**0.917 (0.002)**	**0.911 (0.000)**	0.482 (0.003)	**0.917 (0.002)**	**0.911 (0.000)**	0.482 (0.003)	**0.917 (0.002)**	**0.911 (0.000)**	0.482 (0.003)

Table 18. Summary Statistics for df = banknote; SCAR scheme

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	0.997 (0.002)	0.462 (0.019)	0.624 (0.018)	0.997 (0.002)	0.462 (0.019)	0.624 (0.018)	0.997 (0.002)	0.462 (0.019)	0.624 (0.018)
0.3	clust	0.726 (0.047)	0.665 (0.052)	0.585 (0.063)	0.722 (0.034)	0.651 (0.040)	0.586 (0.044)	0.717 (0.037)	0.634 (0.029)	0.590 (0.033)
0.3	strict-lassclust	0.753 (0.089)	0.635 (0.061)	0.598 (0.095)	0.730 (0.053)	0.639 (0.040)	0.593 (0.053)	0.716 (0.037)	0.632 (0.028)	0.588 (0.033)
0.3	non-strict-lassclust	0.717 (0.050)	0.636 (0.062)	0.579 (0.076)	0.719 (0.036)	0.638 (0.043)	0.573 (0.073)	0.716 (0.037)	0.632 (0.028)	0.588 (0.033)
0.3	lassojoint	**0.999 (0.001)**	**0.985 (0.007)**	**0.983 (0.008)**	**0.999 (0.001)**	**0.985 (0.007)**	**0.983 (0.008)**	**0.999 (0.001)**	**0.985 (0.007)**	**0.983 (0.008)**
0.5	naive	**0.999 (0.001)**	0.707 (0.032)	0.759 (0.024)	**0.999 (0.001)**	0.707 (0.032)	0.759 (0.024)	**0.999 (0.001)**	0.707 (0.032)	0.759 (0.024)
0.5	clust	0.754 (0.011)	0.667 (0.010)	0.575 (0.031)	0.755 (0.012)	0.674 (0.012)	0.601 (0.032)	0.768 (0.010)	0.691 (0.011)	0.648 (0.015)
0.5	strict-lassclust	0.754 (0.011)	0.664 (0.011)	0.571 (0.034)	0.755 (0.011)	0.674 (0.012)	0.601 (0.032)	0.766 (0.010)	0.688 (0.011)	0.644 (0.015)
0.5	non-strict-lassclust	0.754 (0.011)	0.664 (0.011)	0.571 (0.034)	0.755 (0.012)	0.673 (0.012)	0.600 (0.033)	0.766 (0.010)	0.688 (0.011)	0.644 (0.015)
0.5	lassojoint	**0.999 (0.001)**	**0.978 (0.003)**	**0.977 (0.003)**	**0.999 (0.001)**	**0.978 (0.003)**	**0.977 (0.003)**	**0.999 (0.001)**	**0.978 (0.003)**	**0.977 (0.003)**
0.8	naive	0.999 (0.000)	0.925 (0.002)	0.922 (0.003)	0.999 (0.000)	0.925 (0.002)	0.922 (0.003)	0.999 (0.000)	0.925 (0.002)	0.922 (0.003)
0.8	clust	0.788 (0.004)	0.739 (0.003)	0.618 (0.010)	0.798 (0.003)	0.756 (0.011)	0.656 (0.024)	0.832 (0.009)	0.795 (0.007)	0.736 (0.009)
0.8	strict-lassclust	0.787 (0.009)	0.733 (0.015)	0.612 (0.010)	0.799 (0.003)	0.754 (0.012)	0.651 (0.025)	0.832 (0.009)	0.794 (0.005)	0.733 (0.007)
0.8	non-strict-lassclust	0.793 (0.005)	0.740 (0.004)	0.613 (0.010)	0.799 (0.003)	0.754 (0.012)	0.651 (0.025)	0.832 (0.009)	0.794 (0.005)	0.733 (0.007)
0.8	lassojoint	**1.000 (0.000)**	**0.991 (0.004)**	**0.990 (0.005)**	**1.000 (0.000)**	**0.991 (0.004)**	**0.990 (0.005)**	**1.000 (0.000)**	**0.991 (0.004)**	**0.990 (0.005)**

Table 19. Summary Statistics for df = breastc; SCAR scheme

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	0.984 (0.011)	0.348 (0.031)	0.514 (0.034)	0.984 (0.011)	0.348 (0.031)	0.514 (0.034)	0.984 (0.011)	0.348 (0.031)	0.514 (0.034)
0.3	clust	0.982 (0.022)	0.956 (0.013)	0.934 (0.021)	0.982 (0.021)	0.956 (0.014)	0.934 (0.021)	0.982 (0.022)	0.956 (0.014)	0.934 (0.021)
0.3	strict-lassclust	**0.995 (0.003)**	**0.957 (0.013)**	**0.936 (0.020)**	**0.995 (0.003)**	**0.957 (0.013)**	**0.936 (0.020)**	**0.995 (0.003)**	**0.957 (0.013)**	**0.936 (0.020)**
0.3	non-strict-lassclust	**0.995 (0.003)**	**0.957 (0.013)**	**0.936 (0.020)**	**0.995 (0.003)**	**0.957 (0.013)**	**0.936 (0.020)**	**0.995 (0.003)**	**0.957 (0.013)**	**0.936 (0.020)**
0.3	lassojoint	0.982 (0.016)	0.879 (0.176)	0.875 (0.125)	0.983 (0.015)	0.868 (0.191)	0.867 (0.134)	0.983 (0.016)	0.868 (0.191)	0.867 (0.134)
0.5	naive	0.995 (0.003)	0.832 (0.036)	0.811 (0.027)	0.995 (0.003)	0.832 (0.036)	0.811 (0.027)	0.995 (0.003)	0.832 (0.036)	0.811 (0.027)
0.5	clust	**0.997 (0.007)**	0.963 (0.003)	0.946 (0.004)	0.997 (0.007)	0.963 (0.003)	0.946 (0.004)	0.997 (0.009)	0.971 (0.007)	0.958 (0.010)
0.5	strict-lassclust	0.997 (0.001)	**0.966 (0.002)**	**0.950 (0.003)**	**0.998 (0.001)**	**0.969 (0.003)**	**0.954 (0.005)**	**0.998 (0.001)**	**0.973 (0.003)**	**0.961 (0.005)**
0.5	non-strict-lassclust	0.997 (0.001)	**0.966 (0.002)**	**0.950 (0.003)**	**0.998 (0.001)**	**0.969 (0.003)**	**0.954 (0.005)**	**0.998 (0.001)**	**0.973 (0.003)**	**0.961 (0.005)**
0.5	lassojoint	0.995 (0.003)	0.943 (0.025)	0.926 (0.031)	0.995 (0.003)	0.943 (0.025)	0.926 (0.031)	0.995 (0.003)	0.942 (0.025)	0.925 (0.031)
0.8	naive	0.995 (0.001)	0.963 (0.005)	0.951 (0.007)	0.995 (0.001)	0.963 (0.005)	0.951 (0.007)	0.995 (0.001)	0.963 (0.005)	0.951 (0.007)
0.8	clust	**0.996 (0.001)**	0.952 (0.005)	0.930 (0.004)	0.996 (0.000)	0.958 (0.005)	0.938 (0.005)	0.996 (0.000)	0.963 (0.002)	0.946 (0.006)
0.8	strict-lassclust	0.995 (0.000)	0.954 (0.002)	0.933 (0.000)	0.996 (0.000)	0.961 (0.001)	0.943 (0.003)	0.995 (0.000)	0.963 (0.002)	0.946 (0.006)
0.8	non-strict-lassclust	0.995 (0.000)	0.954 (0.002)	0.933 (0.000)	0.996 (0.000)	0.961 (0.001)	0.943 (0.003)	0.995 (0.000)	0.963 (0.002)	0.946 (0.006)
0.8	lassojoint	**0.996 (0.001)**	**0.968 (0.002)**	**0.954 (0.006)**	**0.996 (0.001)**	**0.968 (0.002)**	**0.954 (0.006)**	**0.996 (0.001)**	**0.968 (0.003)**	**0.954 (0.006)**

Table 20. Summary Statistics for df = credit-a; SCAR scheme

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	**0.813 (0.030)**	0.457 (0.029)	0.626 (0.027)	**0.813 (0.030)**	0.457 (0.029)	0.626 (0.027)	**0.813 (0.030)**	0.457 (0.029)	0.626 (0.027)
0.3	clust	0.692 (0.039)	**0.610 (0.039)**	0.366 (0.068)	0.689 (0.036)	**0.611 (0.038)**	0.378 (0.065)	0.684 (0.031)	**0.612 (0.037)**	0.393 (0.061)
0.3	strict-lassclust	0.678 (0.039)	0.601 (0.046)	0.346 (0.090)	0.683 (0.033)	0.607 (0.038)	0.364 (0.072)	0.681 (0.031)	0.609 (0.037)	0.382 (0.064)
0.3	non-strict-lassclust	0.690 (0.039)	0.609 (0.040)	0.364 (0.077)	0.688 (0.037)	0.609 (0.037)	0.369 (0.067)	0.681 (0.031)	0.609 (0.037)	0.382 (0.064)
0.3	lassojoint	0.808 (0.032)	0.570 (0.138)	**0.642 (0.046)**	0.807 (0.033)	0.573 (0.138)	**0.642 (0.046)**	0.809 (0.033)	0.573 (0.136)	**0.642 (0.045)**
0.5	naive	0.822 (0.009)	0.439 (0.018)	0.609 (0.017)	0.822 (0.009)	0.439 (0.018)	0.609 (0.017)	0.822 (0.009)	0.439 (0.018)	0.609 (0.017)
0.5	clust	0.662 (0.016)	0.602 (0.019)	0.332 (0.026)	0.661 (0.015)	**0.612 (0.017)**	0.359 (0.024)	0.658 (0.011)	**0.608 (0.018)**	0.365 (0.022)
0.5	strict-lassclust	0.649 (0.011)	0.599 (0.018)	0.311 (0.035)	0.656 (0.017)	0.609 (0.021)	0.351 (0.029)	0.650 (0.014)	0.606 (0.016)	0.347 (0.027)
0.5	non-strict-lassclust	0.661 (0.020)	**0.602 (0.021)**	0.329 (0.035)	0.657 (0.017)	0.609 (0.021)	0.351 (0.029)	0.650 (0.014)	0.606 (0.016)	0.347 (0.027)
0.5	lassojoint	**0.823 (0.009)**	0.441 (0.017)	**0.612 (0.016)**	**0.823 (0.009)**	0.441 (0.017)	**0.612 (0.016)**	**0.823 (0.009)**	0.441 (0.017)	**0.612 (0.016)**
0.8	naive	0.831 (0.021)	**0.746 (0.028)**	**0.721 (0.033)**	0.831 (0.021)	**0.746 (0.028)**	**0.721 (0.033)**	0.831 (0.021)	**0.746 (0.028)**	**0.721 (0.033)**
0.8	clust	0.674 (0.010)	0.616 (0.007)	0.319 (0.007)	0.674 (0.006)	0.622 (0.014)	0.349 (0.013)	0.673 (0.009)	0.620 (0.010)	0.354 (0.000)
0.8	strict-lassclust	0.671 (0.010)	0.613 (0.012)	0.311 (0.006)	0.668 (0.005)	0.618 (0.005)	0.329 (0.003)	0.667 (0.007)	0.620 (0.010)	0.351 (0.005)
0.8	non-strict-lassclust	0.672 (0.010)	0.609 (0.009)	0.300 (0.001)	0.671 (0.004)	0.615 (0.010)	0.327 (0.000)	0.667 (0.007)	0.620 (0.010)	0.351 (0.005)
0.8	lassojoint	**0.833 (0.021)**	0.414 (0.002)	0.586 (0.002)	**0.834 (0.022)**	0.414 (0.002)	0.586 (0.002)	**0.834 (0.022)**	0.414 (0.002)	0.586 (0.002)

Table 21. Summary Statistics for df = credit-g; SCAR scheme

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	**0.679 (0.042)**	0.299 (0.023)	0.460 (0.027)	**0.679 (0.042)**	0.299 (0.023)	0.460 (0.027)	0.679 (0.042)	0.299 (0.023)	0.460 (0.027)
0.3	clust	0.670 (0.056)	0.587 (0.083)	0.462 (0.220)	0.677 (0.050)	0.554 (0.070)	0.483 (0.170)	**0.687 (0.035)**	0.521 (0.025)	**0.481 (0.092)**
0.3	strict-lassclust	0.642 (0.084)	0.573 (0.088)	**0.544 (0.182)**	0.662 (0.072)	0.548 (0.071)	**0.521 (0.145)**	0.683 (0.047)	0.519 (0.025)	0.481 (0.091)
0.3	non-strict-lassclust	0.667 (0.062)	0.588 (0.086)	0.451 (0.201)	0.672 (0.061)	0.553 (0.073)	0.467 (0.156)	0.683 (0.047)	0.519 (0.025)	0.481 (0.091)
0.3	lassojoint	0.662 (0.063)	**0.596 (0.178)**	0.420 (0.133)	0.661 (0.066)	**0.596 (0.179)**	0.417 (0.138)	0.653 (0.072)	**0.589 (0.181)**	0.437 (0.143)
0.5	naive	0.699 (0.022)	0.306 (0.034)	0.445 (0.013)	0.699 (0.022)	0.306 (0.034)	0.445 (0.013)	0.699 (0.022)	0.306 (0.034)	0.445 (0.013)
0.5	clust	**0.701 (0.021)**	0.657 (0.070)	0.436 (0.154)	**0.702 (0.019)**	0.590 (0.050)	0.497 (0.034)	**0.701 (0.018)**	**0.567 (0.012)**	0.492 (0.011)
0.5	strict-lassclust	0.655 (0.022)	0.639 (0.081)	**0.633 (0.213)**	0.700 (0.019)	0.588 (0.044)	**0.498 (0.034)**	**0.701 (0.018)**	0.567 (0.011)	**0.494 (0.011)**
0.5	non-strict-lassclust	0.697 (0.037)	**0.667 (0.074)**	0.416 (0.148)	0.701 (0.018)	**0.592 (0.056)**	0.469 (0.078)	**0.701 (0.018)**	0.567 (0.011)	**0.494 (0.011)**
0.5	lassojoint	0.674 (0.029)	0.497 (0.218)	0.343 (0.115)	0.665 (0.035)	0.546 (0.214)	0.313 (0.122)	0.679 (0.034)	0.468 (0.214)	0.348 (0.117)
0.8	naive	**0.751 (0.024)**	**0.729 (0.003)**	0.515 (0.001)	**0.751 (0.024)**	**0.729 (0.003)**	0.515 (0.001)	**0.751 (0.024)**	**0.729 (0.003)**	0.515 (0.001)
0.8	clust	0.731 (0.016)	0.719 (0.003)	0.531 (0.006)	0.731 (0.016)	0.718 (0.004)	0.530 (0.009)	0.731 (0.016)	0.718 (0.004)	0.530 (0.009)
0.8	strict-lassclust	0.731 (0.016)	0.724 (0.008)	**0.535 (0.012)**	0.731 (0.016)	0.725 (0.007)	**0.539 (0.012)**	0.731 (0.016)	0.728 (0.004)	**0.543 (0.011)**
0.8	non-strict-lassclust	0.731 (0.016)	0.724 (0.008)	**0.535 (0.012)**	0.731 (0.016)	0.725 (0.007)	**0.539 (0.012)**	0.731 (0.016)	0.728 (0.004)	**0.543 (0.011)**
0.8	lassojoint	0.741 (0.018)	0.718 (0.013)	0.208 (0.053)	0.741 (0.018)	0.718 (0.013)	0.208 (0.053)	0.741 (0.018)	0.718 (0.013)	0.208 (0.053)

Table 22. Summary Statistics for df = dhfr; SCAR scheme

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	0.592 (0.074)	0.534 (0.083)	0.559 (0.068)	0.592 (0.074)	0.534 (0.083)	0.559 (0.068)	0.592 (0.074)	0.534 (0.083)	0.559 (0.068)
0.3	clust	0.557 (0.048)	0.571 (0.063)	0.451 (0.053)	0.559 (0.075)	0.577 (0.083)	0.451 (0.067)	0.551 (0.038)	0.561 (0.053)	0.446 (0.051)
0.3	strict-lassclust	0.592 (0.074)	0.648 (0.040)	**0.709 (0.119)**	0.630 (0.066)	**0.634 (0.031)**	**0.672 (0.139)**	0.617 (0.079)	0.579 (0.076)	0.474 (0.097)
0.3	non-strict-lassclust	0.755 (0.071)	0.647 (0.096)	0.590 (0.053)	0.766 (0.076)	0.610 (0.117)	0.540 (0.184)	0.617 (0.079)	0.579 (0.076)	0.474 (0.097)
0.3	lassojoint	**0.866 (0.047)**	**0.659 (0.174)**	0.675 (0.095)	**0.832 (0.122)**	0.585 (0.200)	0.640 (0.097)	**0.881 (0.046)**	**0.633 (0.187)**	**0.665 (0.092)**
0.5	naive	0.606 (0.061)	0.572 (0.051)	0.562 (0.055)	0.606 (0.061)	0.572 (0.051)	0.562 (0.055)	0.606 (0.061)	0.572 (0.051)	0.562 (0.055)
0.5	clust	0.603 (0.052)	**0.643 (0.039)**	0.520 (0.037)	0.509 (0.082)	0.576 (0.040)	0.457 (0.014)	0.509 (0.082)	0.576 (0.040)	0.457 (0.014)
0.5	strict-lassclust	0.577 (0.035)	0.625 (0.008)	**0.740 (0.092)**	0.621 (0.013)	**0.632 (0.066)**	0.468 (0.011)	0.621 (0.016)	0.593 (0.030)	0.478 (0.019)
0.5	non-strict-lassclust	0.823 (0.071)	0.453 (0.108)	0.553 (0.037)	0.625 (0.008)	0.577 (0.021)	0.469 (0.005)	0.621 (0.016)	0.593 (0.030)	0.478 (0.019)
0.5	lassojoint	**0.886 (0.028)**	0.636 (0.120)	0.653 (0.060)	**0.892 (0.024)**	0.596 (0.104)	**0.628 (0.054)**	**0.897 (0.021)**	**0.606 (0.107)**	**0.645 (0.057)**
0.8	naive	0.724 (0.000)	0.711 (0.000)	0.659 (0.000)	0.724 (0.000)	0.711 (0.000)	0.659 (0.000)	0.724 (0.000)	0.711 (0.000)	0.659 (0.000)
0.8	clust	0.680 (0.000)	0.701 (0.000)	0.540 (0.000)	0.650 (0.000)	0.680 (0.000)	0.492 (0.000)	0.626 (0.000)	0.639 (0.000)	0.462 (0.000)
0.8	strict-lassclust	0.680 (0.000)	**0.773 (0.000)**	0.542 (0.000)	0.674 (0.000)	0.773 (0.000)	0.560 (0.000)	0.675 (0.000)	0.794 (0.000)	0.643 (0.000)
0.8	non-strict-lassclust	0.683 (0.000)	0.691 (0.000)	0.559 (0.000)	0.681 (0.000)	0.670 (0.000)	0.543 (0.000)	0.675 (0.000)	0.794 (0.000)	0.643 (0.000)
0.8	lassojoint	**0.925 (0.000)**	0.742 (0.000)	**0.713 (0.000)**	**0.920 (0.000)**	**0.907 (0.000)**	**0.862 (0.000)**	**0.920 (0.000)**	**0.907 (0.000)**	**0.862 (0.000)**

Table 23. Summary Statistics for df = diabetes; SCAR scheme

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	0.794 (0.032)	0.352 (0.030)	0.516 (0.032)	0.794 (0.032)	0.352 (0.030)	0.516 (0.032)	0.794 (0.032)	0.352 (0.030)	0.516 (0.032)
0.3	clust	0.723 (0.035)	0.679 (0.026)	0.451 (0.086)	0.719 (0.032)	0.679 (0.025)	0.472 (0.042)	0.717 (0.029)	0.678 (0.025)	0.477 (0.042)
0.3	strict-lassclust	0.709 (0.037)	0.672 (0.028)	0.431 (0.124)	0.711 (0.032)	0.674 (0.027)	0.445 (0.090)	0.712 (0.029)	0.676 (0.026)	0.473 (0.041)
0.3	non-strict-lassclust	0.720 (0.036)	0.679 (0.028)	0.474 (0.073)	0.715 (0.032)	0.677 (0.026)	0.474 (0.043)	0.712 (0.029)	0.676 (0.026)	0.473 (0.041)
0.3	lassojoint	**0.796 (0.043)**	**0.682 (0.128)**	**0.593 (0.063)**	**0.800 (0.032)**	**0.685 (0.124)**	**0.597 (0.061)**	**0.800 (0.033)**	**0.681 (0.131)**	**0.592 (0.069)**
0.5	naive	**0.819 (0.002)**	0.506 (0.027)	0.613 (0.026)	**0.819 (0.002)**	0.506 (0.027)	0.613 (0.026)	**0.819 (0.002)**	0.506 (0.027)	0.613 (0.026)
0.5	clust	0.724 (0.011)	0.648 (0.021)	0.437 (0.011)	0.725 (0.010)	0.651 (0.022)	0.447 (0.012)	0.723 (0.011)	0.651 (0.020)	0.440 (0.013)
0.5	strict-lassclust	0.709 (0.018)	0.650 (0.019)	0.440 (0.013)	0.709 (0.018)	0.651 (0.019)	0.446 (0.012)	0.708 (0.017)	0.651 (0.019)	0.446 (0.012)
0.5	non-strict-lassclust	0.709 (0.018)	0.650 (0.019)	0.439 (0.012)	0.709 (0.018)	0.651 (0.019)	0.446 (0.012)	0.708 (0.017)	0.651 (0.019)	0.446 (0.012)
0.5	lassojoint	0.812 (0.006)	**0.717 (0.070)**	**0.669 (0.033)**	0.812 (0.006)	**0.717 (0.070)**	**0.669 (0.033)**	0.813 (0.005)	**0.717 (0.070)**	**0.669 (0.033)**
0.8	naive	0.842 (0.006)	**0.776 (0.009)**	**0.713 (0.007)**	0.842 (0.006)	**0.776 (0.009)**	**0.713 (0.007)**	0.842 (0.006)	**0.776 (0.009)**	**0.713 (0.007)**
0.8	clust	0.768 (0.007)	0.669 (0.006)	0.396 (0.006)	0.768 (0.007)	0.669 (0.006)	0.396 (0.006)	0.764 (0.006)	0.664 (0.005)	0.393 (0.005)
0.8	strict-lassclust	0.764 (0.007)	0.664 (0.005)	0.383 (0.005)	0.765 (0.007)	0.669 (0.005)	0.395 (0.008)	0.760 (0.005)	0.664 (0.004)	0.392 (0.007)
0.8	non-strict-lassclust	0.764 (0.007)	0.664 (0.005)	0.383 (0.005)	0.765 (0.007)	0.669 (0.006)	0.395 (0.008)	0.760 (0.005)	0.664 (0.004)	0.392 (0.007)
0.8	lassojoint	**0.846 (0.008)**	0.761 (0.002)	0.655 (0.003)	**0.845 (0.008)**	0.761 (0.002)	0.651 (0.004)	**0.845 (0.008)**	0.761 (0.002)	0.651 (0.004)

Table 24. Summary Statistics for df = spambase; SCAR scheme

calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	0.936 (0.010)	0.397 (0.011)	0.565 (0.011)	0.936 (0.010)	0.397 (0.011)	0.565 (0.011)	0.936 (0.010)	0.397 (0.011)	0.565 (0.011)
0.3	clust	0.876 (0.009)	0.821 (0.009)	0.756 (0.012)	0.876 (0.009)	0.820 (0.009)	0.756 (0.012)	0.876 (0.009)	0.820 (0.009)	0.757 (0.012)
0.3	strict-lassclust	0.874 (0.009)	0.819 (0.009)	0.753 (0.012)	0.874 (0.009)	0.819 (0.009)	0.754 (0.012)	0.874 (0.009)	0.818 (0.009)	0.755 (0.012)
0.3	non-strict-lassclust	0.874 (0.009)	0.819 (0.009)	0.753 (0.012)	0.874 (0.009)	0.819 (0.009)	0.754 (0.012)	0.874 (0.009)	0.818 (0.009)	0.755 (0.012)
0.3	lassojoint	**0.946 (0.011)**	**0.872 (0.100)**	**0.852 (0.063)**	**0.946 (0.011)**	**0.864 (0.119)**	**0.848 (0.073)**	**0.946 (0.012)**	**0.867 (0.110)**	**0.849 (0.069)**
0.5	naive	0.952 (0.005)	0.550 (0.028)	0.626 (0.013)	0.952 (0.005)	0.550 (0.028)	0.626 (0.013)	0.952 (0.005)	0.550 (0.028)	0.626 (0.013)
0.5	clust	0.898 (0.007)	0.838 (0.009)	0.765 (0.011)	0.898 (0.007)	0.838 (0.010)	0.768 (0.012)	0.897 (0.006)	0.838 (0.009)	0.770 (0.009)
0.5	strict-lassclust	0.894 (0.008)	0.837 (0.009)	0.763 (0.011)	0.893 (0.007)	0.837 (0.007)	0.765 (0.008)	0.893 (0.006)	0.836 (0.007)	0.766 (0.008)
0.5	non-strict-lassclust	0.894 (0.008)	0.837 (0.009)	0.763 (0.011)	0.893 (0.007)	0.837 (0.007)	0.765 (0.008)	0.893 (0.006)	0.836 (0.007)	0.766 (0.008)
0.5	lassojoint	**0.963 (0.006)**	**0.915 (0.008)**	**0.888 (0.010)**	**0.960 (0.009)**	**0.849 (0.152)**	**0.843 (0.105)**	**0.963 (0.003)**	**0.915 (0.005)**	**0.888 (0.007)**
0.8	naive	**0.951 (0.001)**	**0.897 (0.004)**	**0.871 (0.005)**	**0.951 (0.001)**	**0.897 (0.004)**	**0.871 (0.005)**	**0.951 (0.001)**	**0.897 (0.004)**	**0.871 (0.005)**
0.8	clust	0.911 (0.002)	0.833 (0.001)	0.751 (0.001)	0.912 (0.001)	0.836 (0.000)	0.759 (0.001)	0.912 (0.001)	0.841 (0.000)	0.767 (0.001)
0.8	strict-lassclust	0.904 (0.001)	0.822 (0.002)	0.733 (0.003)	0.906 (0.001)	0.829 (0.000)	0.746 (0.001)	0.907 (0.001)	0.833 (0.000)	0.755 (0.001)
0.8	non-strict-lassclust	0.904 (0.001)	0.822 (0.002)	0.733 (0.003)	0.906 (0.001)	0.831 (0.001)	0.750 (0.002)	0.907 (0.001)	0.833 (0.000)	0.755 (0.001)
0.8	lassojoint	0.936 (0.000)	0.386 (0.001)	0.555 (0.001)	0.936 (0.000)	0.386 (0.001)	0.555 (0.001)	0.936 (0.000)	0.386 (0.001)	0.555 (0.001)

Table 25. Summary Statistics for df = wdbc; SCAR scheme

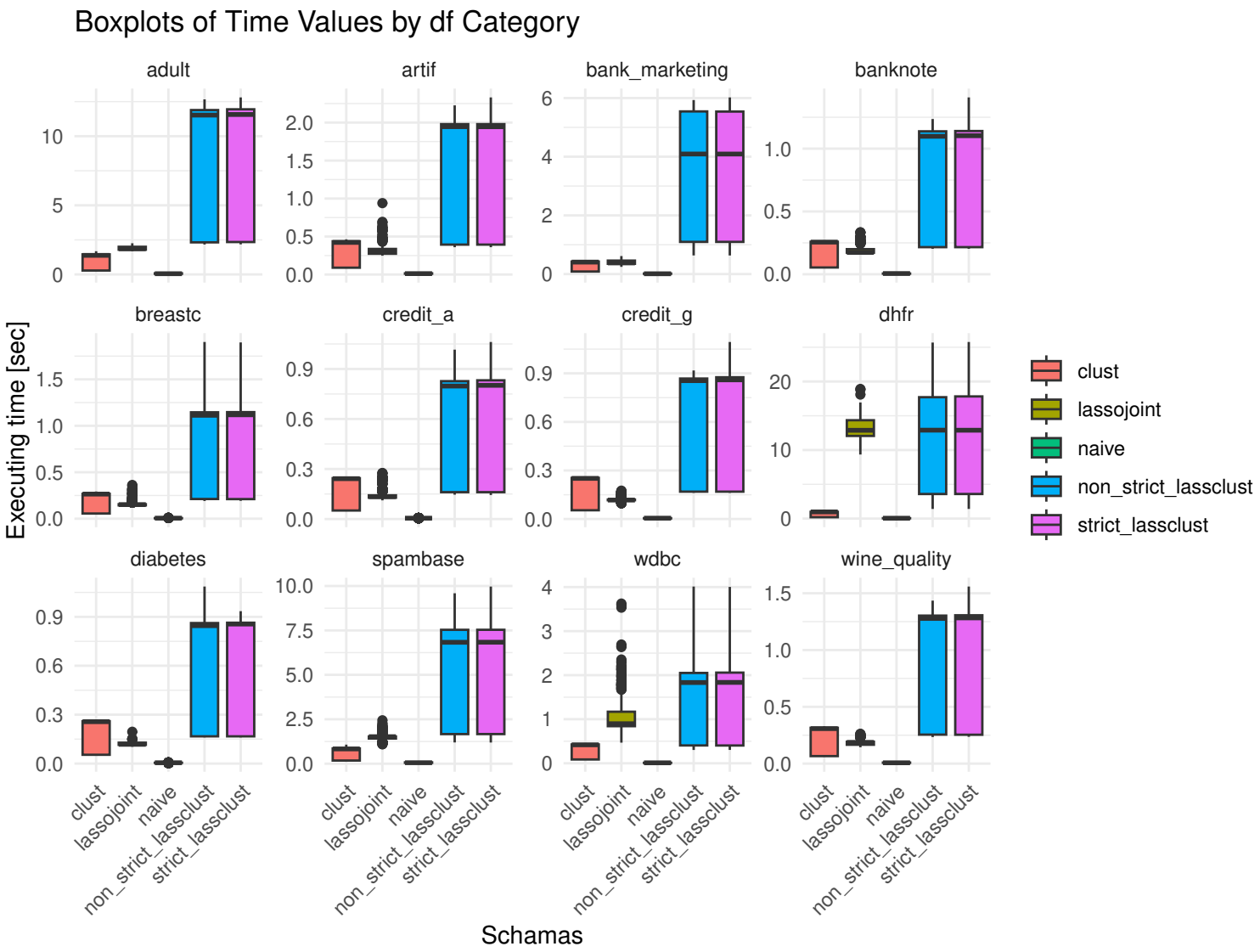
calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	0.959 (0.022)	0.435 (0.037)	0.566 (0.029)	0.959 (0.022)	0.435 (0.037)	0.566 (0.029)	0.959 (0.022)	0.435 (0.037)	0.566 (0.029)
0.3	clust	0.919 (0.022)	0.907 (0.020)	0.865 (0.030)	0.918 (0.022)	0.908 (0.020)	0.866 (0.030)	0.919 (0.022)	0.908 (0.020)	0.867 (0.030)
0.3	strict-lassclust	0.971 (0.013)	0.881 (0.057)	0.802 (0.144)	0.970 (0.014)	0.878 (0.062)	0.794 (0.157)	**0.973 (0.010)**	**0.916 (0.017)**	**0.878 (0.027)**
0.3	non-strict-lassclust	0.973 (0.010)	**0.916 (0.021)**	**0.882 (0.028)**	0.973 (0.010)	**0.915 (0.022)**	**0.883 (0.028)**	**0.973 (0.010)**	**0.916 (0.017)**	**0.878 (0.027)**
0.3	lassojoint	**0.973 (0.021)**	0.883 (0.111)	0.867 (0.082)	**0.974 (0.020)**	0.880 (0.121)	0.866 (0.088)	0.971 (0.023)	0.872 (0.125)	0.858 (0.091)
0.5	naive	**0.991 (0.004)**	0.649 (0.021)	0.648 (0.024)	**0.991 (0.004)**	0.649 (0.021)	0.648 (0.024)	**0.991 (0.004)**	0.649 (0.021)	0.648 (0.024)
0.5	clust	0.930 (0.015)	0.923 (0.007)	0.873 (0.009)	0.934 (0.020)	**0.938 (0.010)**	**0.898 (0.014)**	0.940 (0.022)	0.924 (0.011)	0.874 (0.019)
0.5	strict-lassclust	0.983 (0.004)	0.870 (0.067)	0.710 (0.233)	0.979 (0.005)	0.874 (0.054)	0.754 (0.128)	0.986 (0.002)	**0.938 (0.004)**	**0.900 (0.007)**
0.5	non-strict-lassclust	0.984 (0.004)	**0.927 (0.006)**	**0.888 (0.007)**	0.987 (0.005)	0.932 (0.008)	0.898 (0.012)	0.986 (0.002)	**0.938 (0.004)**	**0.900 (0.007)**
0.5	lassojoint	0.990 (0.007)	0.917 (0.033)	0.887 (0.036)	0.990 (0.007)	0.917 (0.033)	0.887 (0.036)	0.990 (0.007)	0.918 (0.034)	0.890 (0.038)
0.8	naive	**0.991 (0.003)**	0.920 (0.011)	0.894 (0.019)	0.991 (0.003)	0.920 (0.011)	0.894 (0.019)	0.991 (0.003)	0.920 (0.011)	0.894 (0.019)
0.8	clust	0.924 (0.007)	0.916 (0.003)	0.867 (0.018)	0.931 (0.006)	0.926 (0.011)	0.883 (0.028)	0.945 (0.008)	0.941 (0.000)	0.911 (0.005)
0.8	strict-lassclust	0.952 (0.011)	0.817 (0.057)	0.645 (0.117)	0.984 (0.006)	0.768 (0.116)	0.451 (0.312)	0.982 (0.002)	0.946 (0.009)	0.915 (0.021)
0.8	non-strict-lassclust	0.984 (0.005)	0.908 (0.003)	0.878 (0.001)	0.981 (0.003)	0.754 (0.150)	0.756 (0.117)	0.982 (0.002)	0.946 (0.009)	0.915 (0.021)
0.8	lassojoint	0.990 (0.001)	**0.974 (0.002)**	**0.963 (0.006)**	**0.994 (0.002)**	**0.972 (0.005)**	**0.959 (0.011)**	**0.992 (0.000)**	**0.974 (0.002)**	**0.963 (0.006)**

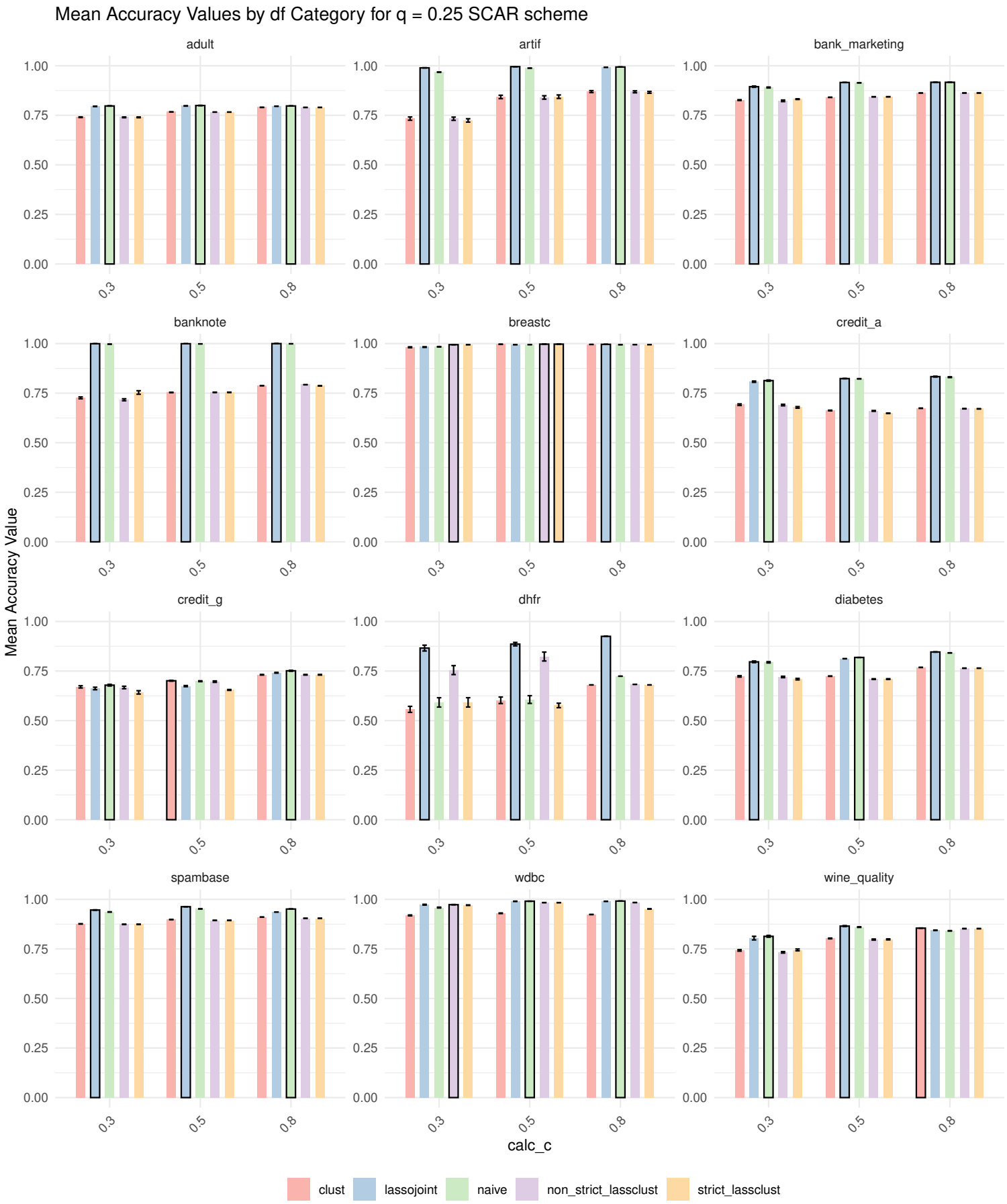
Table 26. Summary Statistics for df = wine-quality; SCAR scheme

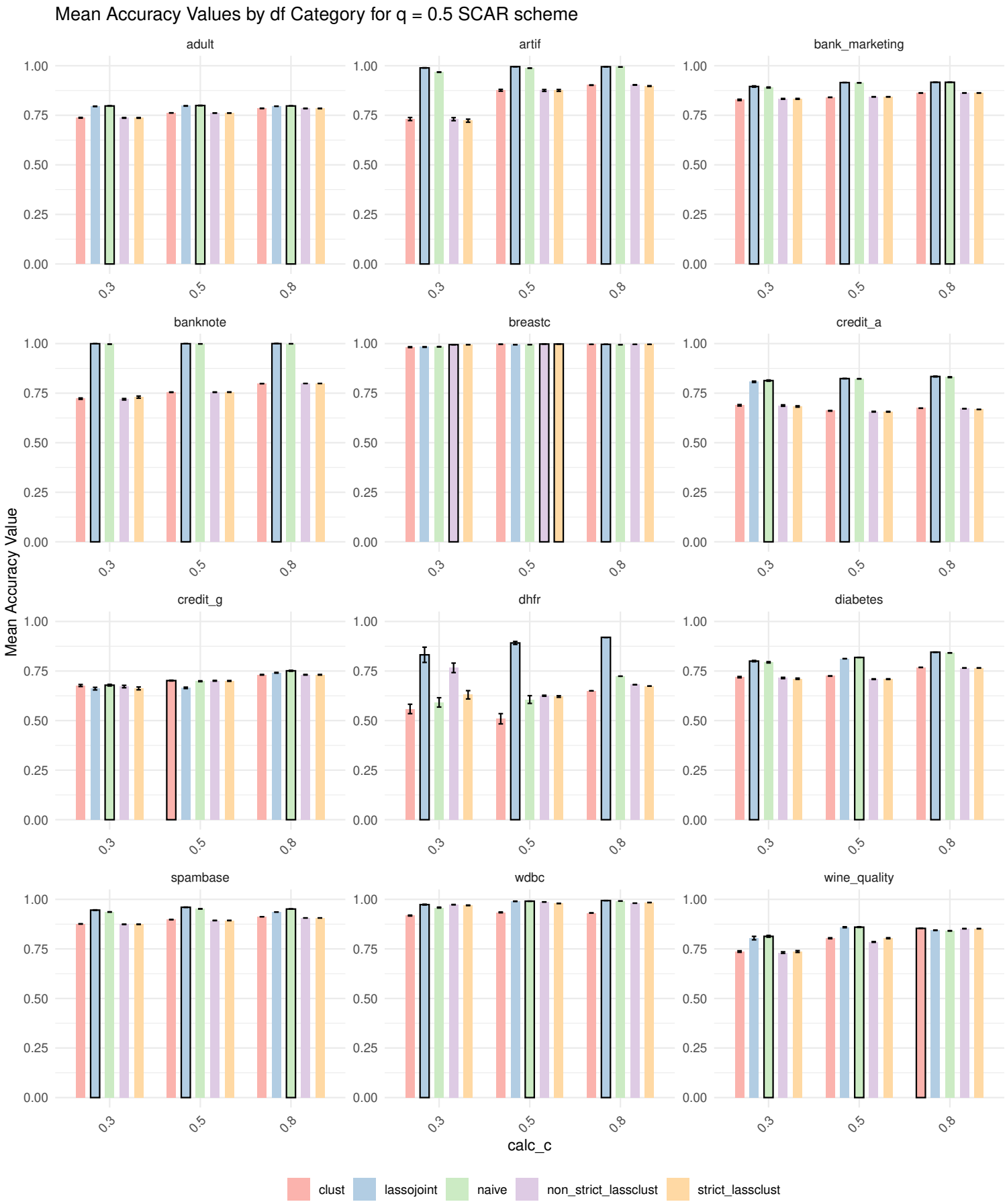
calc-c	method	0.25 auc	0.25 acc	0.25 f1	0.5 auc	0.5 acc	0.5 f1	1 auc	1 acc	1 f1
0.3	naive	**0.814 (0.046)**	0.137 (0.013)	0.238 (0.019)	**0.814 (0.046)**	0.137 (0.013)	0.238 (0.019)	**0.814 (0.046)**	0.137 (0.013)	0.238 (0.019)
0.3	clust	0.742 (0.037)	**0.765 (0.061)**	0.347 (0.115)	0.737 (0.039)	**0.740 (0.055)**	0.382 (0.085)	0.737 (0.030)	**0.724 (0.062)**	**0.390 (0.072)**
0.3	strict-lassclust	0.745 (0.046)	0.752 (0.091)	0.353 (0.131)	0.737 (0.045)	0.737 (0.071)	**0.387 (0.099)**	0.735 (0.029)	0.723 (0.061)	0.386 (0.071)
0.3	non-strict-lassclust	0.733 (0.037)	0.753 (0.070)	0.359 (0.120)	0.732 (0.039)	0.730 (0.071)	0.382 (0.100)	0.735 (0.029)	0.723 (0.061)	0.386 (0.071)
0.3	lassojoint	0.804 (0.090)	0.685 (0.293)	**0.381 (0.155)**	0.805 (0.086)	0.724 (0.266)	0.385 (0.155)	0.808 (0.080)	0.698 (0.281)	0.380 (0.157)
0.5	naive	0.860 (0.017)	0.349 (0.041)	0.294 (0.011)	**0.860 (0.017)**	0.349 (0.041)	0.294 (0.011)	**0.860 (0.017)**	0.349 (0.041)	0.294 (0.011)
0.5	clust	0.803 (0.019)	0.824 (0.009)	**0.455 (0.053)**	0.804 (0.024)	0.830 (0.022)	0.415 (0.047)	0.799 (0.021)	**0.812 (0.011)**	**0.462 (0.016)**
0.5	strict-lassclust	0.798 (0.024)	0.809 (0.017)	0.423 (0.123)	0.804 (0.025)	0.814 (0.009)	**0.451 (0.014)**	0.796 (0.022)	0.811 (0.011)	0.447 (0.018)
0.5	non-strict-lassclust	0.797 (0.028)	0.815 (0.031)	0.451 (0.032)	0.785 (0.019)	0.799 (0.023)	0.414 (0.024)	0.796 (0.022)	0.811 (0.011)	0.447 (0.018)
0.5	lassojoint	**0.865 (0.024)**	**0.853 (0.085)**	0.443 (0.060)	0.859 (0.024)	**0.868 (0.013)**	0.335 (0.092)	0.859 (0.023)	0.791 (0.229)	0.303 (0.086)
0.8	naive	0.841 (0.012)	0.846 (0.004)	0.415 (0.075)	0.841 (0.012)	0.846 (0.004)	0.415 (0.075)	0.841 (0.012)	0.846 (0.004)	0.415 (0.075)
0.8	clust	**0.855 (0.008)**	**0.869 (0.006)**	**0.466 (0.039)**	**0.854 (0.009)**	**0.871 (0.007)**	**0.475 (0.028)**	**0.852 (0.009)**	0.868 (0.010)	0.456 (0.030)
0.8	strict-lassclust	0.852 (0.009)	0.868 (0.009)	0.457 (0.023)	0.852 (0.009)	0.869 (0.010)	0.462 (0.028)	0.851 (0.010)	**0.872 (0.011)**	**0.457 (0.040)**
0.8	non-strict-lassclust	0.852 (0.009)	0.868 (0.009)	0.453 (0.043)	0.852 (0.009)	0.869 (0.010)	0.461 (0.034)	0.851 (0.010)	**0.872 (0.011)**	**0.457 (0.040)**
0.8	lassojoint	0.844 (0.011)	0.862 (0.005)	0.348 (0.079)	0.844 (0.011)	0.862 (0.005)	0.346 (0.082)	0.844 (0.011)	0.862 (0.005)	0.346 (0.082)

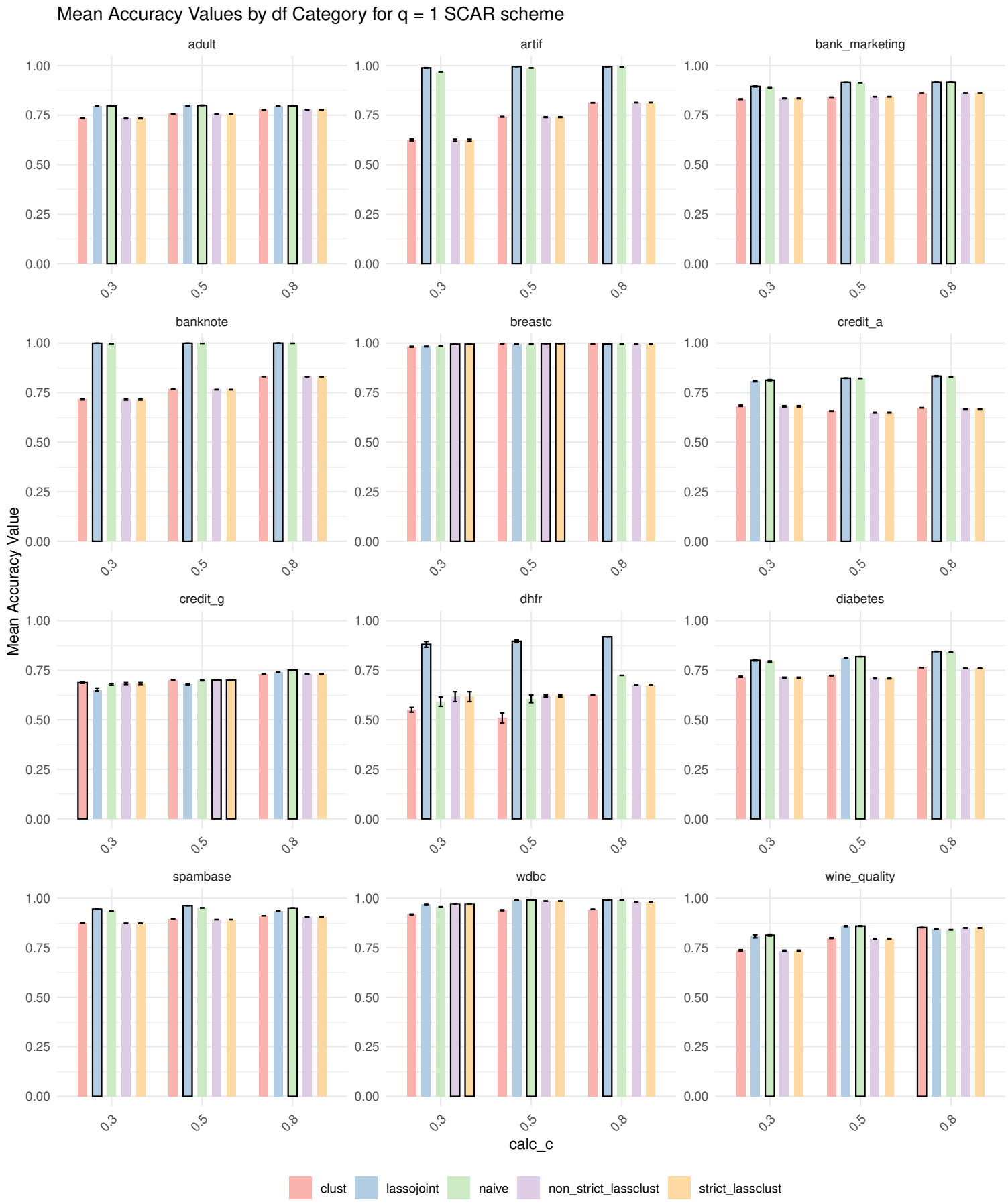
2.2 Graphs

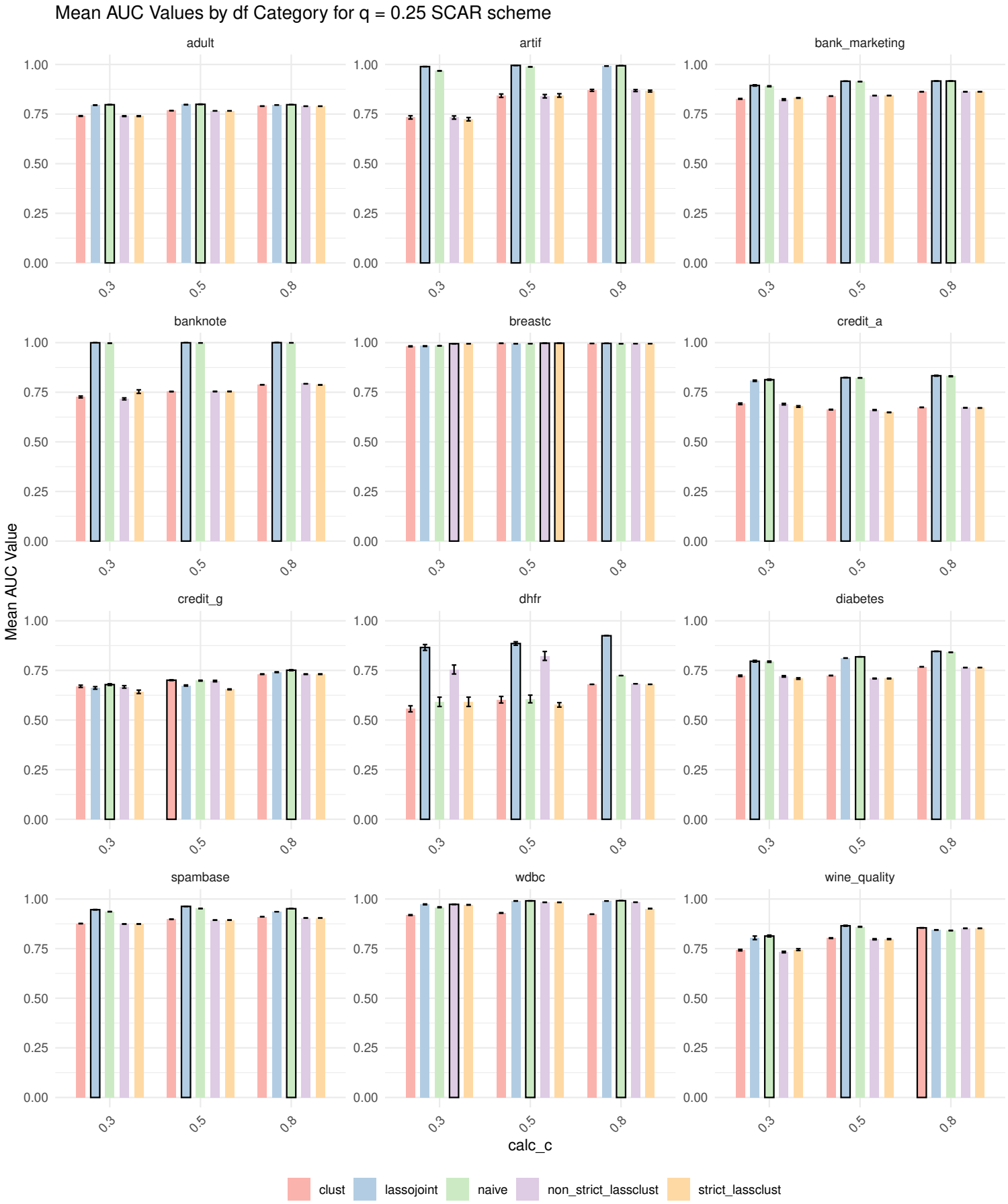
In this section we present all barplots for classification metrics for our methods. In addition, we present boxplots of the execution times of the algorithms used.

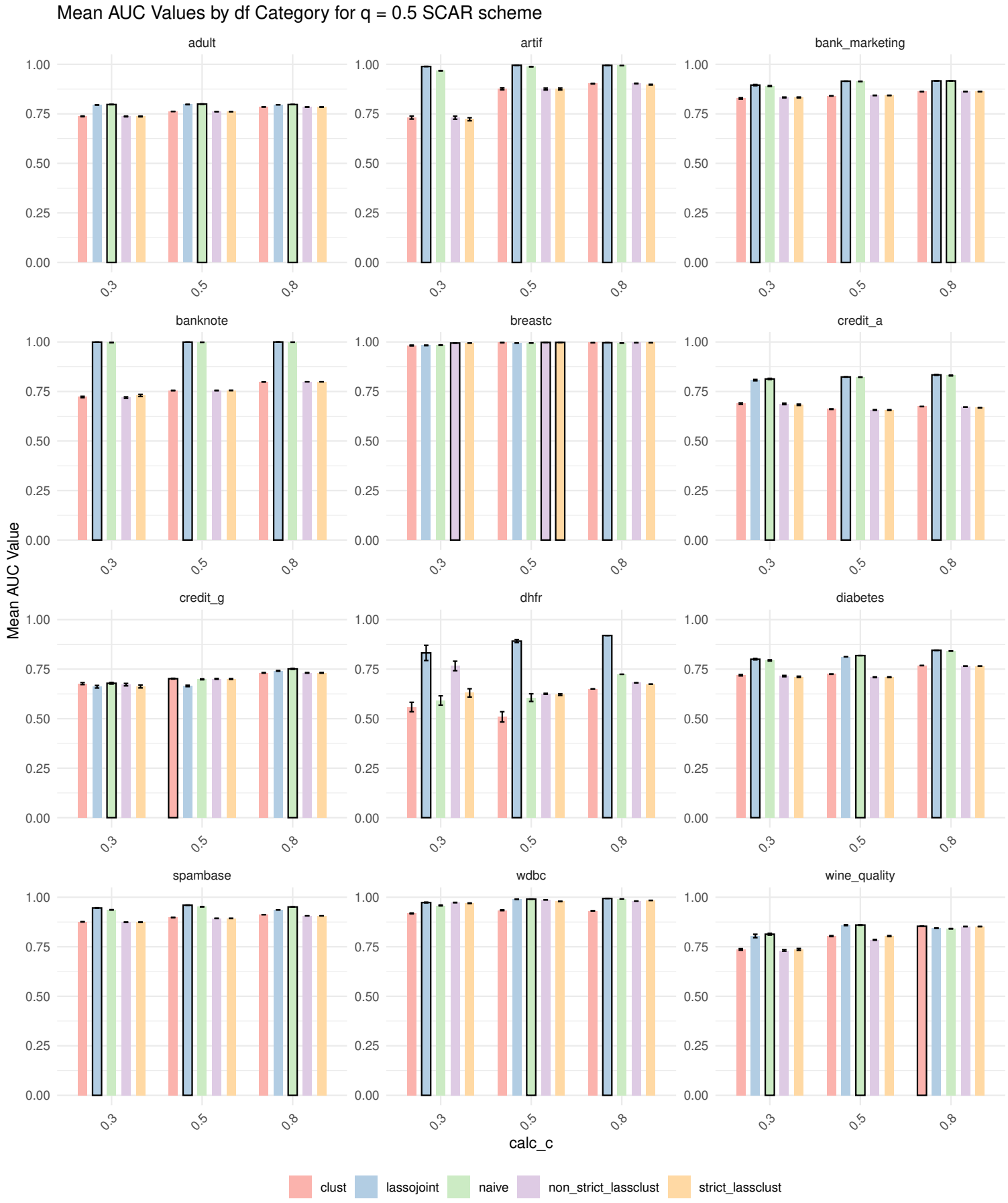

Boxplots of Time Values by df Category
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Executing time [sec]
Schamas
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust

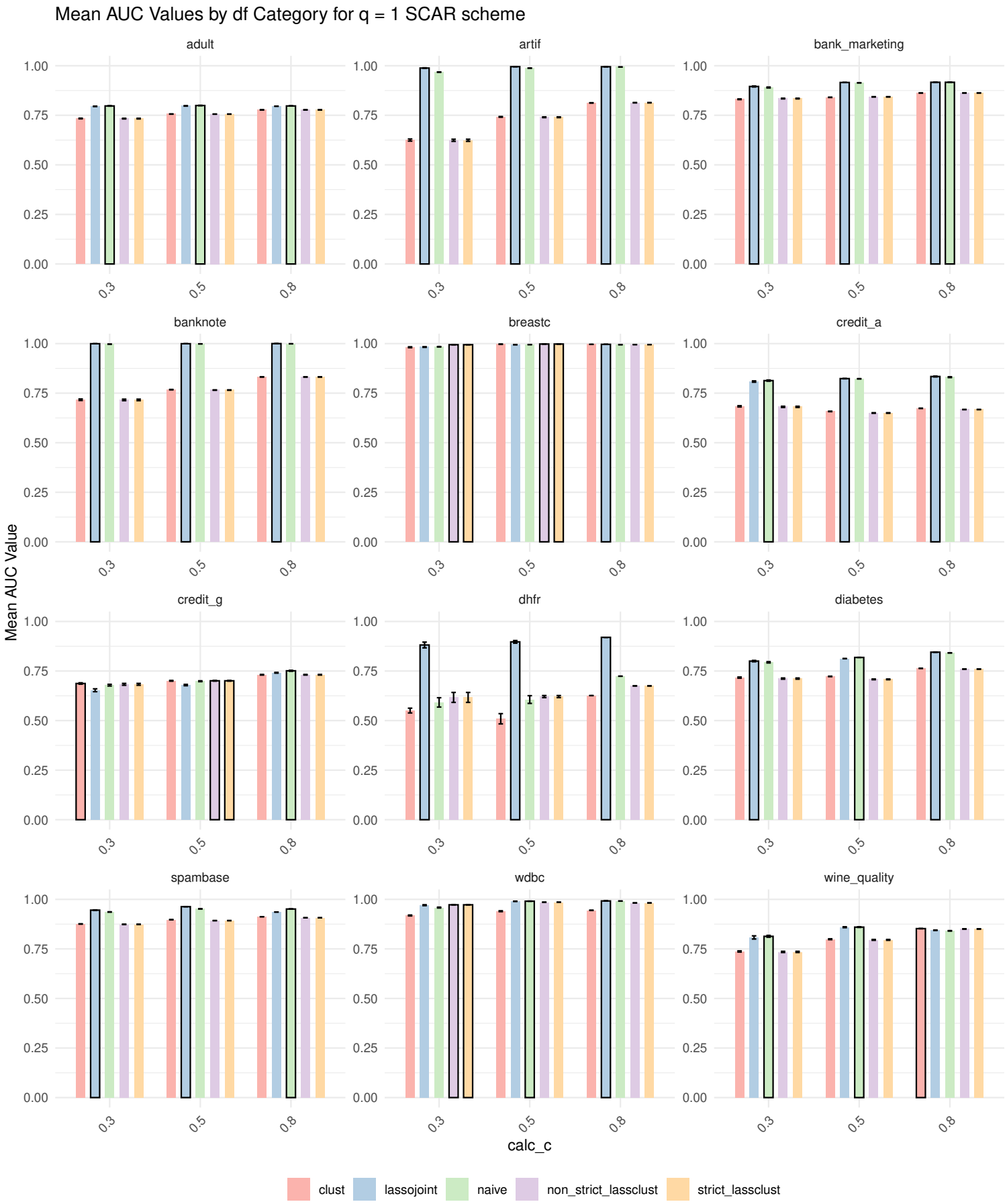

Mean Accuracy Values by df Category for q = 0.25 SCAR scheme
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean Accuracy Value
calc_c
0.00
0.25
0.50
0.75
1.00
0.3
0.5
0.8
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust

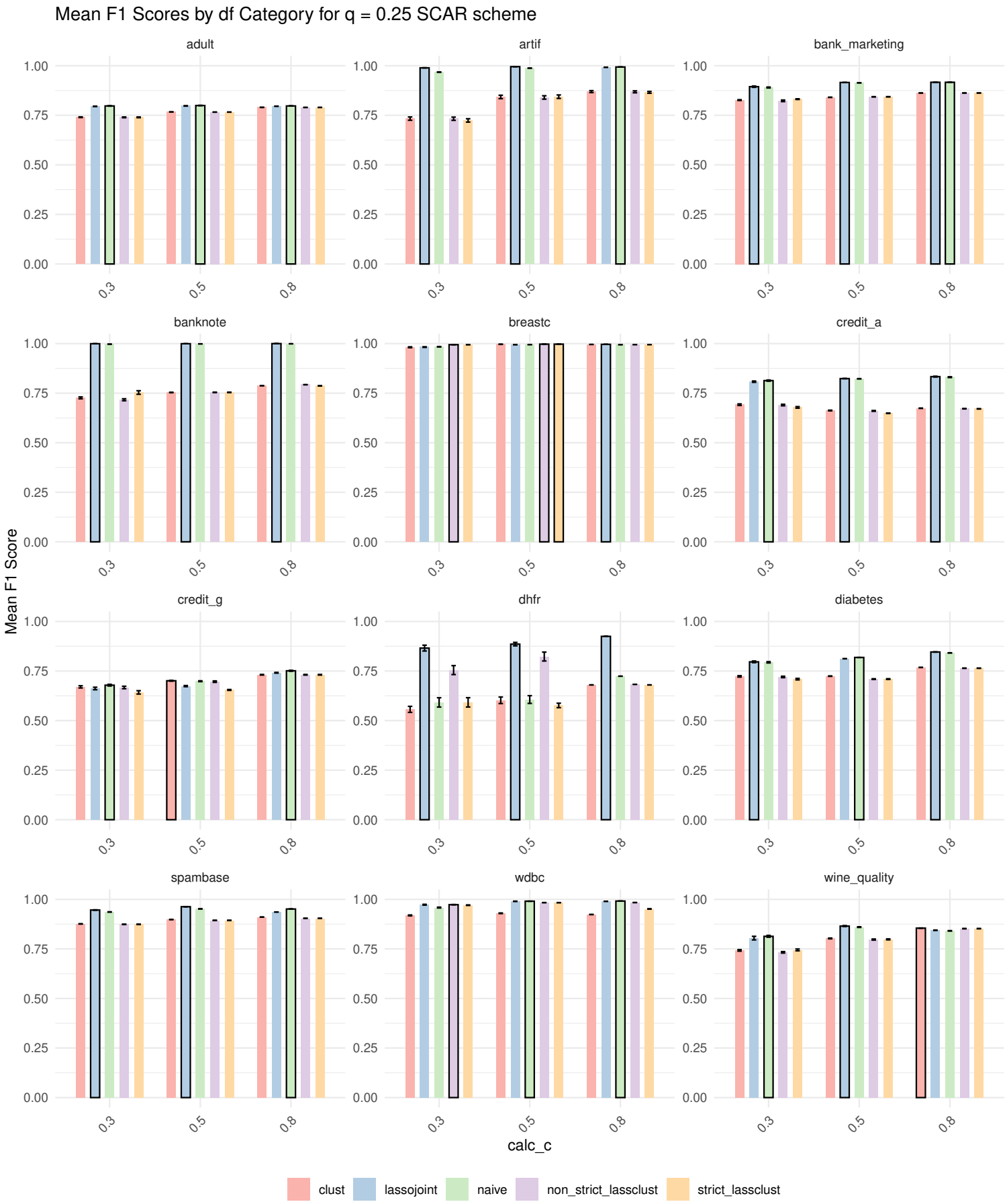

Mean Accuracy Values by df Category for q = 0.5 SCAR scheme
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean Accuracy Value
1.00
0.75
0.50
0.25
0.00
0.3
0.5
0.8
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust

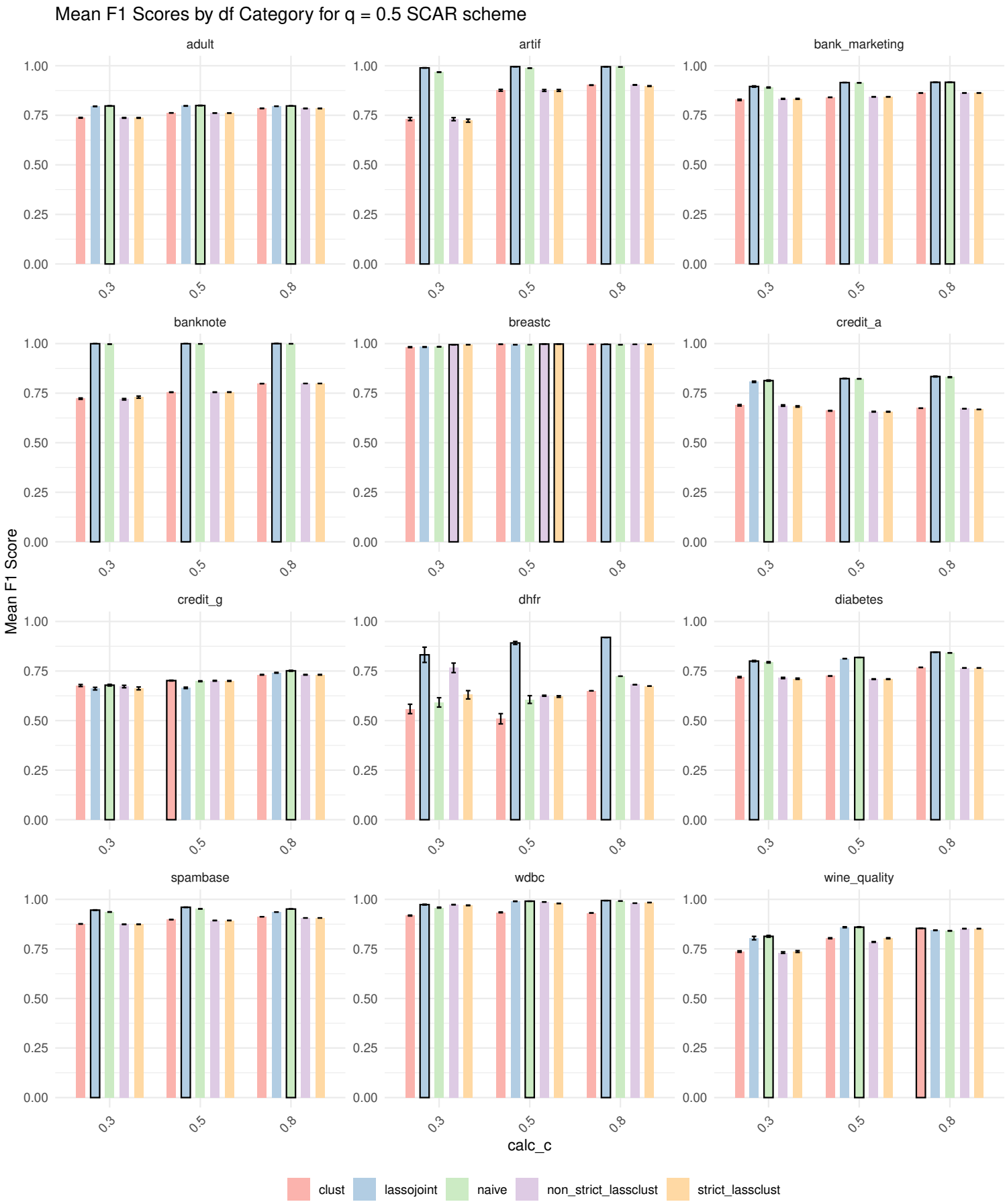

Mean Accuracy Values by df Category for q = 1 SCAR scheme
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean Accuracy Value
1.00
0.75
0.50
0.25
0.00
0.3
0.5
0.8
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust

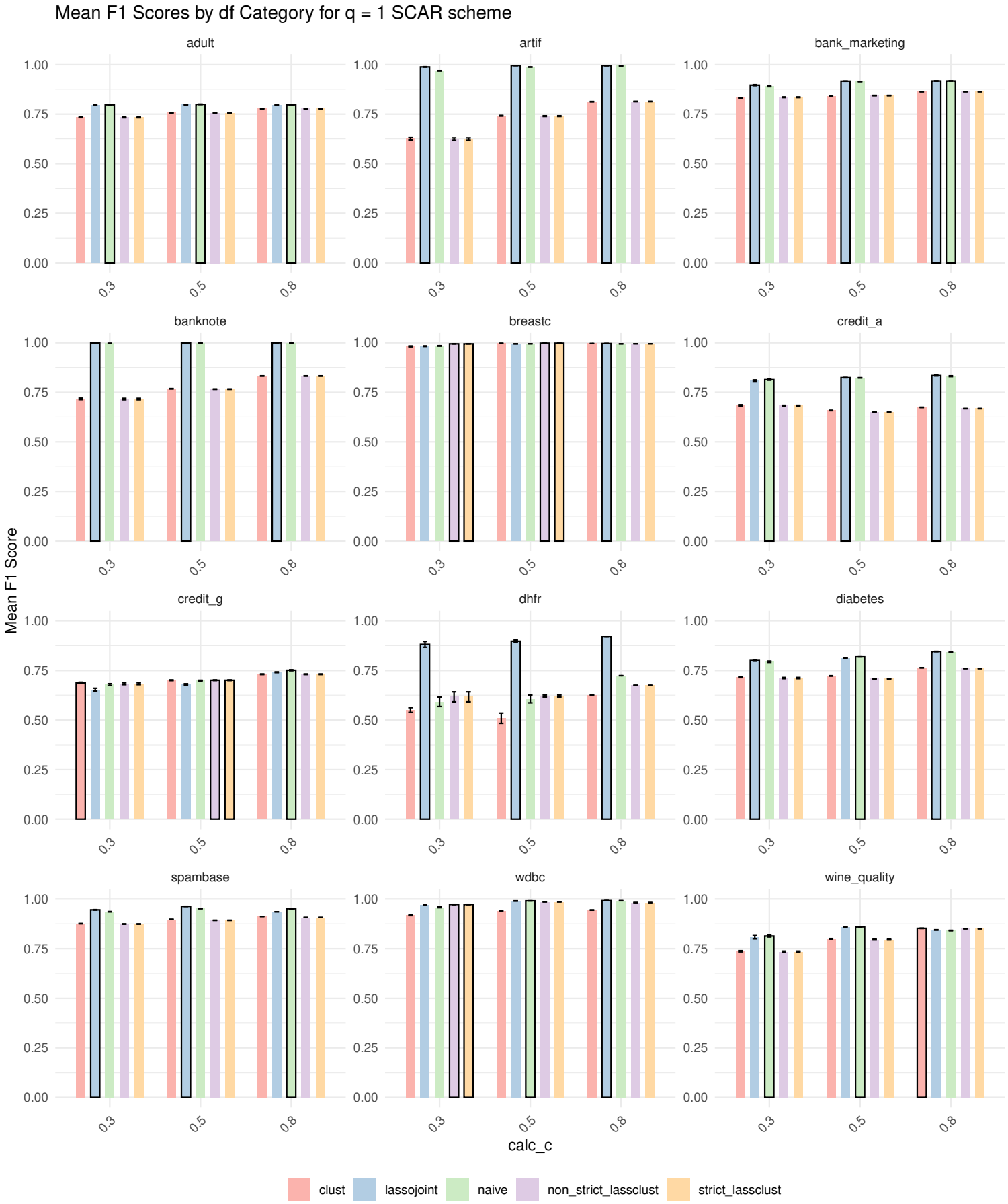

Mean AUC Values by df Category for q = 0.25 SCAR scheme
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean AUC Value
1.00
0.75
0.50
0.25
0.00
0.3
0.5
0.8
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust


Mean AUC Values by df Category for q = 0.5 SCAR scheme
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean AUC Value
1.00
0.75
0.50
0.25
0.00
0.3
0.5
0.8
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust


Mean AUC Values by df Category for q = 1 SCAR scheme
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean AUC Value
1.00
0.75
0.50
0.25
0.00
0.3
0.5
0.8
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust


Mean F1 Scores by df Category for q = 0.25 SCAR scheme
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean F1 Score
1.00
0.75
0.50
0.25
0.00
0.3
0.5
0.8
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust


Mean F1 Scores by df Category for q = 0.5 SCAR scheme
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean F1 Score
1.00
0.75
0.50
0.25
0.00
0.3
0.5
0.8
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust


Mean F1 Scores by df Category for q = 1 SCAR scheme
adult
artif
bank_marketing
banknote
breastc
credit_a
credit_g
dhfr
diabetes
spambase
wdbc
wine_quality
Mean F1 Score
1.00
0.75
0.50
0.25
0.00
0.3
0.5
0.8
calc_c
clust
lassojoint
naive
non_strict_lassclust
strict_lassclust

3 Description of the Function `non_scar_labelling.mvc`

Given a dataset $\mathbf{D}$ with columns representing features and a target variable Y, the function `non_scar_labelling.mvc` performs the following steps:

1. Filter out columns Y:

$$\mathbf{D}_{\text{filtered}} = \mathbf{D} \setminus \{Y\}$$

2. Calculate variance for each column:

$$\text{var}_i = \text{Var}(X_i) \quad \text{for each feature } X_i \in \mathbf{D}_{\text{filtered}}$$

3. Sort columns by decreasing variance:

$$\text{sorted\_variances} = \text{sort}(\{\text{var}_i\}, \text{decreasing} = \text{TRUE})$$

4. Select top n_{vars} columns with highest variance:

$$\mathbf{T} = \{X_i \mid \text{var}_i \in \text{sorted\_variances}[1 : n_{\text{vars}}]\}$$

5. Create temporary dataframe with new column defined as rank of sums of selected column, where rank rn=1 for the smallest value and rank rn=n for the largest value and n is the number of rows in our dataset :

$$\mathbf{D}_{\text{temp}} = \mathbf{D} \cup \left\{ \text{rn} = \text{row\_number}\left(\sum_{X_i \in \mathbf{T}} X_i \right) \right\}$$

6. Define binary variables S_{02}, S_{05}, and S_{08}:

$$S_{02} = \begin{cases} 1 & \text{if } \frac{\text{rn}}{n} < 0.2 \text{ and } Y = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$S_{05} = \begin{cases} 1 & \text{if } \frac{\text{rn}}{n} < 0.5 \text{ and } Y = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$S_{08} = \begin{cases} 1 & \text{if } \frac{\text{rn}}{n} < 0.8 \text{ and } Y = 1 \\ 0 & \text{otherwise} \end{cases}$$

7. Calculate c_{calc} values:

$$c_{\text{calc02}} = \frac{\sum S_{02}}{\sum Y}, \quad c_{\text{calc05}} = \frac{\sum S_{05}}{\sum Y}, \quad c_{\text{calc08}} = \frac{\sum S_{08}}{\sum Y}$$

$$\mathbf{c}_{\text{calc}} = \{c_{\text{calc02}}, c_{\text{calc05}}, c_{\text{calc08}}\}$$

8. Fit a linear model to estimate target c_{calc}:

$$c_{\text{calc}} \sim \alpha + \beta \times x, \quad \text{where } x \in \{0.2, 0.5, 0.8\}$$

$$\text{rn\_frac} = \frac{\text{target\_c\_calc} - \alpha}{\beta}$$

9. Create the final dataframe with the new column S:

$$S = \begin{cases} 1 & \text{if } \frac{\text{rn}}{n} < \text{rn\_frac and } Y = 1 \\ 0 & \text{otherwise} \end{cases}$$

The final dataframe $\mathbf{D1} = \mathbf{D} \cup S$ is returned with the new column S representing the target variable's adjusted binary label.